REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE 12/18/01	3. REPORT TYPE AND DATES COVERED Final, ~~2/1/98-2/1/01~~ 01 Mar 98 - 28 Feb 01

4. TITLE AND SUBTITLE

Interferometric Digital Imaging

5. FUNDING NUMBERS

ARO Contract 38310-PH
DAAG55-98-1-0039

6. AUTHOR(S)

David J. Brady, David C. Munson and Eric Michielssen

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Beckman Institute
University of Illinois
405 N. Mathews
Urbana, IL 61801

8. PERFORMING ORGANIZATION REPORT NUMBER

RECEIVED
JAN 0 9 2002
By

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Dr. David Skatrud
U. S. Army Research Office
P. O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

38310.6-PH

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)

The Illinois Interferometric Imaging Initiative (4Is) demonstrated that aspects of conventional focal imaging, coherence sensing, holography, spectroscopy and tomography could be integrated to obtain revolutionary new functionalities in digital imaging systems. Novel functions demonstrated included multidimensional imaging, phase and polarization sensitive imaging, simultaneous spatial and spectral analysis and nonlocal basis vector sampling. The 4Is initiative provides strong evidence that the "pixels-to-pupils ratio" problem of distributed sensor systems can be addressed by designing sensors to detect target specific physical primitives. These primitives can be adaptively or interactively identified and isolated. Initiatives to develop scan-free 3D microscopes, spatio-spectral biosensors, advanced tracking telescopes and cooperative sensor arrays have resulted from the 4Is initiative.

14. SUBJECT TERMS	15. NUMBER OF PAGES whole bunch
	16. PRICE CODE

| 17. SECURITY CLASSIFICATION OF REPORT unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT unclassified	20. LIMITATION OF ABSTRACT ul

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

20020125 279

Fitzpatrick Center for Photonics and Communications Systems
Duke University
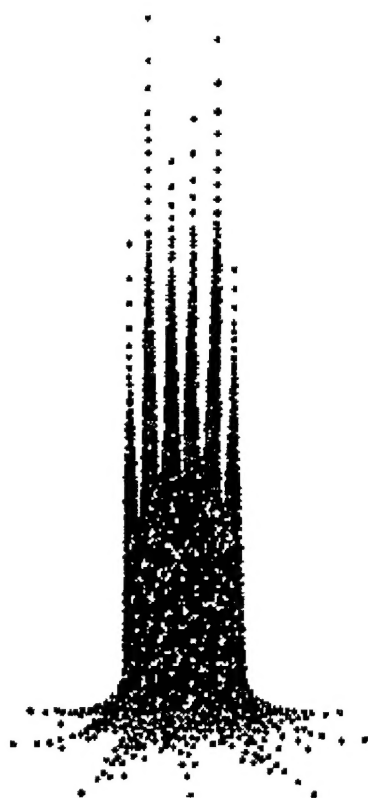
Interferometric Digital Imaging

Final Progress Report

ARO Contract 38310-PH
DAAG55-98-1-0039

December, 2001

Table of Contents

Program Overview

"Interferometric Digital Imaging," a DARPA DSO ACMP program funded under ARO management as ARO Contract 38310-PH DAAG55-98-1-0039 examined the fundamental limits of information transfer through optical apertures. Conventional imaging systems perform both a data acquisition and a data processing role. The interferometric digital imaging program focused on the question: "Can the data acquisition rate of an aperture be increased by changing or reducing the conventional data processing role?" Three years of investigation have shown that there exist interesting and technologically important situations in which the answer to this question is yes.

This project exposed opportunities for mixed-mode optical and opto-electronic processing, *ad hoc* networking and aggressive digital processing on optical and infrared imaging systems. The project demonstrated near-term opportunities, including:

- The use of the sampling field sensor (SFS) for process control and materials characterization in optical and electronic manufacturing,
- The use of the SFS in multi-wavelength mode for ranging and 3D imaging,
- The use of the SFS for wavefront characterization,
- The use of the rotational shear interferometer for 3D microscopy and the potential for scan-free interferometric 3D microscopy,
- The use of rotational shear interferometers for high speed spatio-spectral tracking,
- The potential for direct physical measurement of source projections on nonlocal bases,
- The potential use extended depth-of-field interferometer arrays for spatio-spectral aerosol analysis,
- The use of coherence mode decomposition for distortion correction and imaging through turbulence,
- The potential for global transformations, rather than correlation interferometery, for background-free coherence characterization.
- The potential use of extended depth of field interferometric sensors for multiple access free space optical communications.

These opportunities suggest that digital interferometery will produce significant changes in optical and infrared imaging systems over the full range from microscopy to telescopy.

Conventionally the data processing role of an imaging system is to form a physical isomorphism between a source space and a measurement space. For example, a pixel on a camera focal plane corresponds to the field on a source plane. Imaging systems that do not form such isomorphisms use inversion algorithms to estimate source state parameters from measurements that depend nonlocally on the source state. We refer to these systems as *multiplex* imagers. The primary advantage of multiplexing is the freedom it gives the designer to incorporate global system performance measures into system design. There is only one isomorphic transformation between the source state and the measurement space, there are infinitely many multiplex transformations. Multiplex system design consists of choosing the transformation that achieves the best rate and accuracy of source parameter

estimation within physical and deployment constraints. Because the multiplex design methodology combines physical constraints and logical analysis goals the ACMP community has adopted the phrase "integrated sensing and processing" (ISP) to describe it. If we were to start again in 1998, "Integrated sensing and processing in multiplex digital imaging" would be the best name for the program reported here.

Inteferometric Digital Imaging was the core project of the Illinois Interferometric Imaging Initiative (4Is). 4Is began with an exploration in the summer of 1995 into the possibility of lensless imaging by sampling and digital processing. This exploration led to the development of the sampling field sensor in 1996. Development of the sampling field sensor was initially supported by the University of Illinois and the Packard Foundation. Following the construction of an initial prototype the SFS project was described to Drs. L. N. Durvasula and Dennis Healy in the summer of 1997, which led to the start of DARPA support in February 1998. In this report, "this program," "Interferometric Digital Imaging program," and "4Is" are treated as synonymous.

Initial work on rotational shear interferometers had already started by the beginning of the DARPA program and it was already clear to the 4Is team that the most significant results of this program would be the delineation of processing and sensing capabilities in multiplex systems rather than the details of specific demonstrations. A high priority was placed on demonstrations, however, as existence proofs of the fact that new functionalities would emerge from unconventional imagers.

4Is directly spawned four successors: the Duke Integrated Sensing and Processing group (DISP), the DISP Information Spaces Project (also DISP), Distant Focus Corporation and Phase Optics, Incorporated. The integrated sensing and processing group includes two DARPA supported initiatives, one funded under the ACMP/DSO program and managed through ARO in computational microscopy for scan-free 3D imaging and projection on nonlocal bases and one in under the PWASSP/MTO program and managed through AFOSR. The information spaces project has been supported by AFOSR and through substantial direct investments from Duke University and the University of Illinois. Distant Focus Corporation completed a DARPA funded tomographic ground sensor project under subcontract to the University of Illinois and is currently developing a wireless sensing and processing platform for heterogeneous multiplex infrared tracking and target recognition. Distant Focus is working with Raytheon West on applications of this technology to flying object identification and tracking. For more information, see www.distantfocus.com Phase Optics licensed sampling field sensor technology from the University of Illinois and received seed financing for commercializing this product in the context of semiconductor manufacturing process control. See www.phaseoptics.com for more information.

We report here (1) research results from optical systems, system capability analyses, data analysis algorithms and methodologies, (2) technology transfer results aimed at introducing the results of this program in deployable systems, (3) opportunities for further development exposed by this program and (4) personnel and publications resulting from this program

Research Results

The Sensor Problem

The Illinois Interferometric Imaging Initiative (4Is) focused on information transfer on optical fields. The process of information transfer on radiating fields consists of

1. data encoding in a source state,
2. coupling from the source state to the field,
3. propagation of the field,
4. modulation of the field,
5. detection of the field and
6. estimation of the source state from the detected field.

Steps 1-3 may be considered as a physical layer beyond the reach of sensor designers. (Perturbation of this layer changes the source state). The *sensor problem* consists of the design of steps 4-6 to maximize measures of system performance. 4Is considered whether or not radical new approaches to steps 4-6 might be taken in optical systems.

For example, coherent scalar radiation from a transparent source in a homogeneous space is described by a superposition of Huygens wavelets or, more generally, by a linear transformation of the form

$$g(\mathbf{r}') = \int f(\mathbf{r}) h(\mathbf{r},\mathbf{r}') d^3\mathbf{r} \tag{1}$$

where $f(\mathbf{r})$ is the primary source density, $g(\mathbf{r}')$ is the radiated field and $h(\mathbf{r},\mathbf{r}')$ is the impulse response. $f(\mathbf{r})$ is the source state. Eqn. (1) expresses the process of information transfer through step 3.

The conventional goal of optical system design is to develop optics capable of inverting Eqn. (1). One places components in the propagation space of the field to implement a transformation $\tilde{h}(\mathbf{r}',\mathbf{r}'')$ on $g(\mathbf{r}')$. The transformed field is

$$\tilde{g}(\mathbf{r}') = \int g(\mathbf{r}') \tilde{h}(\mathbf{r}',\mathbf{r}'') d^2 r' \tag{2}$$

In the ideal conventional system, $\int h(\mathbf{r},\mathbf{r}') \tilde{h}(\mathbf{r}',\mathbf{r}'') d^2 r' = \delta(\mathbf{r} - \mathbf{r}'')$, in which case the measured transformed field is isomorphic to the source state.

4Is challenged the implicit and explicit assumptions of conventional sensor design. These assumptions include:

1. The possibility of isomorphic transformations. The class of sources that can be isomorphically mapped by physical transformation from the source state to the

5

sensor state is a small subset of the class of all possible source states. In general, only two dimensional monochromatic sources can be isomorphically mapped.

2. The desirability of isomorphic transformations. Even in cases where isomorphic transformation is possible, such transformations do not achieve the best system performance against all performance metrics.

3. The nature of source coding and inversion. One may often be interested in distributed properties of sources, "As in how red is the source." Abstraction from locally isomorphic maps may not be efficient in estimation of these generalized source states.

4. The nature of fields and media. In optical systems the coherent electro-magnetic field is generally not directly measureable. The intensity field is measurable, but is not necessarily fully descriptive of the source state. Dimensionality and field transformations play critical roles in analysis of the field states. The change in dimensionality from Eqn. (1) to Eqn. (2) reflects the fact that sources are generally three dimensional while coherent fields are fully characterized from 2D boundary conditions. More general field models are needed to describe the relationship between general source states and radiated fields.

4Is examined these assumptions by

1. Exploring the limits of direct sampling of the electro-magnetic field. If the field could be directly measured, all possible transformations and information from the field could be obtained by digital processing. 4Is used the sampling field sensor to directly measure coherent fields.

2. Exploring the limits of direct sampling of coherence fields. In most optical systems the electro-magnetic field is not measureable. The radiation field is fully and fundamentally characterized by coherence measures. 4Is used the rotational shear interferometer and the astigmatic coherence sensor to directly characterize the coherence field, which then could be digitally processed to estimate the source state.

3. Exploring integration of physical and tomographic inversion of radiation measures. 4Is developed novel algorithms and applied existing algorithms to novel situations to show that analog and digital processing could be integrated to analyze multidimensional sources.

4. Exploring nonlocal transformations for efficient source state estimation.

5. Exploring the physical limits of multiplexing.

Results from these investigations are briefly summarized in the following sections. In each case, substantially more detailed discussions are included as attachments in the Appendix. Our goal here is to state as simply and briefly as possible the significance of each component.

Sampling Field Sensors

As discussed in detail in the documents included in the appendix, the sampling field sensor (SFS) is a wavefront sensor. Wavefront sensors, such as the Shack-Hartman,

shearing interferometer and phase diversity interferometer, are used to measure weak wavefront distortions for adaptive optical systems. The goal is to build systems to conventional isomorphic maps by sensing and correcting distortions. At the level of strongly inhomogeneous wavefronts, holograms may also be considered as wavefront sensors. The goal of the SFS is to digitally sense wavefronts as complex as those sensed by holographic systems. The SFS may be considered as a self-referencing electronic hologram.

The SFS uses interference between multiple samples of a wavefront to estimate the phase and amplitude of the wavefront. One assumes that the wavefront can be sampled at Nyquist determined frequencies or better and that fan-out to interfere samples can be implemented. The SFS uses local differential fan-out, although there global fan-out may be desirable for robust wavefront estimation.

When the 4Is program was launched a preliminary design and test of the SFS had been completed. During the early stages of the 4Is program extensive finite element electromagnetic evaluation of diffractive SFS fan-out systems showed that cross-talk would be a serious issue in all purely diffractive systems. In response to this difficulty, we designed birefringent sampling and fan-out systems that dramatically reduced cross-talk. The SFS has demonstrated extremely competitive phase estimation capabilities, as described in the appendix.

Several promising extensions to SFS systems have yet to be explored. Most interesting, preliminary analysis indicates that SFS sensors could obtain substantial range resolution by sequential multi-color wavefront sensing. Effectively, one would form an electronic hologram of a scene at multiple colors. Assuming that the spectral dependence of the scene is weak, the 3D structure, including range, of the scene could be abstracted from the sequence of wavefronts. Birefringent SFS systems are also capable of estimating polarization images of scenes, which can be used as physical primitives in source analysis.

The University of Illinois has a patent pending on birefringent SFS technology. This patent has been licensed to Phase Optics, inc., which is pursuing commercialization based on semiconductor fabrication process control.

Rotational Shear Interferometers

The assumed association between the source state and the radiation field is one of the deepest challenges of optical sensing. The assumption that the source is linearly related to the coherent field is may be accurate for coherently illuminated sources, but is not otherwise satisfactory. Many potential relationships exist between the source and the radiation state. The simplest assumption for self-luminous and ambiently illuminated sources is to a linear relationship between the source state and a coherence field. This relationship takes the form

$$g(\mathbf{r'},\mathbf{r''}) = \int f(\mathbf{r})h^*(\mathbf{r},\mathbf{r''})h(\mathbf{r},\mathbf{r'})d^3\mathbf{r} \qquad (3)$$

$h(\mathbf{r},\mathbf{r}')$ is the coherent impulse response. Eqn. (3) is obtained from Eqn. (1) by assuming that the source is an incoherent radiator. In the context of optical coherence theory Eqn. (3) is called the Hopkins integral.

Like the field itself, $g(\mathbf{r}',\mathbf{r}'')$ is not directly measureable. It is possible, however, to design interferometers that measure easily invertible functions of $g(\mathbf{r}',\mathbf{r}'')$. The simplest cases are two-point correlators, which measure samples of the form $g(\mathbf{r}',\mathbf{r}') + g(\mathbf{r}'',\mathbf{r}'') + g(\mathbf{r}',\mathbf{r}'') + g(\mathbf{r}'',\mathbf{r}')$. More complex transformations are described below in the ACS section.

Our general goal in sensor system design is to place objects in the radiation space of the field such that data in the field is revealed on propagation. The need for global transformations on propagation is the most significant aspect of the shift from field sensors to coherence sensors. In the case of the SFS, all of the information in the coherent field can be obtained from local sampling and low bandwidth differential measurements are sufficient to fully characterize the field. For fields characterized by coherence functions, associations between fields from widely separated points may be information bearing. To obtain this information the optical system must be capable of interfering widely separated points in the homogeneous radiation space.

The rotational shear interferometer (RSI) is a particularly significant testbed for coherence field characterization because each pixel in the RSI sensor plane measures $g(\mathbf{r}',\mathbf{r}'')$ for a unique separation $\Delta\mathbf{r} = \mathbf{r}' - \mathbf{r}''$.

4Is achieved three major results from RSI-based imaging systems:

1. We showed analytically and experimentally that coherence imaging could be applied to Fresnel zone and three dimensional imaging. To our knowledge, the 1999 Applied Optics manuscript included here as an appendix, was the first derivation of coordinate systems for Fresnel zone analysis.
2. We showed, in the Science paper included as an appendix, that infinite depth of field coherence images could be combined with computed tomography to capture 3D models of sources. In continuing work, we believe that this result will lead to scan-free 3D optical microscopes.
3. We showed that RSI-based imagers could filter out projections of multidimensional source data cubes without first capturing isomorphic data. This result is described theoretically below in the "efficient source state estimation" section and experimentally in Jason Gallicchio's Master of Science dissertation, which is included as an appendix.

Astigmatic Coherence Sensors

While full characterization of the field radiated by an incoherent source implies full characterization of field coherence measures, such characterization need not imply direct measurement of the coherence functions by two-point correlation. As part of the 4Is initiative, we developed a novel coherence sensor based on transformations of the coherence functions by refractive systems. The astigmatic coherence sensor consists of a set of spinning cylindrical lenses. The system measures transformations of the mutual intensity in the input aperture as a function of position in the output aperture, longitudinal position of the output aperture and astigmatisim of the refractive system.

The ACS is particularly noteworthy in two respects:

1. It captures the mutual intensity without the bias noise associated with two-point correlations and
2. It captures the four-dimensional mutual intensity, which enables analysis of generally partially coherent fields.

Detailed analysis of noise issues in multiplex imaging systems requires assumptions regarding the nature of the source. Different sensor systems are adapted to different source classes. For example, focal imaging systems are well adapted to sources with high entropy across a plane and RSI's tend to be well adapted to sparse longitudinally distributed sources. The ACS is an existence proof of the possibility of wholly new classes of sensor system. Ultimately the best sensors are adapted to the source it observes, detailed comparative research into the adaptive range and ideal source characteristics of different sensors is an important topic for continuing study.

We wrote two manuscripts describing the function of the ACS, one on the basic process of coherence measures from refractive transformations and one showing distortion correction by four dimensional coherence analysis. These manuscripts are included as appendices.

Tomographic Analysis of Optical Images

Tomography originally referred to slice selection in 3D object analysis by coded illumination and sensing. With the advent of computed tomography, however, tomography has referred to increasingly broad classes of multidimensional source reconstruction from linear projections. Relationships and boundaries between physical data capture and digital reconstruction have become interesting research topics in many modalities, including x-rays, optics, ultra-sound and radar.

The boundary between physical and digital processing for source estimation is particularly interesting in optics. Optical systems have generally relied on very simple relationships between sources and fields (the field is usually assumed to be simply

9

proportional to the source) and on physical processing for source estimation. Tomographic analysis allows more sophisticated relationships between source and field. For example, tomographic algorithms automatically account for the fact that the field at a given source point may include contributions from other source points. Optical systems may have previously attempted to account for this fact by diffractive deconvolution, but robust well-posed inversion has been relatively rare.

4Is very quickly encountered interesting questions on the boundary between focal imaging and tomography. Tomography is by nature a multiplex imaging technique. Focal systems assume isomorphism. On the one hand, it is not possible to gather tomographic projections in optical systems ray by ray, on the other hand focal systems are not well-posed for projective deconvolution.

After demonstrating tomographic projection from pinhole and focal apertures in the papers included as appendices here, the 4Is project addressed the issue of trade-offs between aperture processing and incoherent tomographic processing in the RSI-based Science article mentioned in the previous section. We also developed a detailed model for the tomographic patch response associated with using conventional CT algorithms on opaque objects. The patch response was the subject of A. J. Johnson's Master of Science thesis and will appear in a future manuscript.

We also considered the efficiency and data requirements for tomographic processing, as indicated by articles (16) and (17) in the publications section.

General issues regarding trade-offs between tomographic processing of incoherent aperture data and coherent analog in-aperture transformations remain unresolved at the time of this report. It now seems clear that computational imaging systems consisting of large arrays of low resolution apertures may achieve comparable or even superior source estimation performance in some applications, but full analysis of this issue awaits further work.

Efficient Source State Estimation
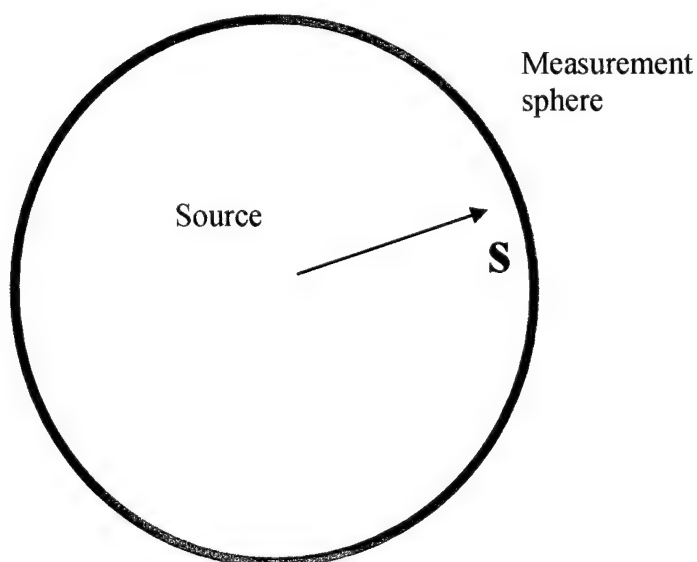
4Is demonstrated by existence that many physical primitives, such as the basic size, radiation pattern and spectrum, of sources may be abstracted more efficiently than digital analysis of isomorphic sensor data. Detailed comparative analysis of different sensors for different source analysis tasks awaits further work. This section presents a short

explanation of the potential for feature specific sensor systems.

Consider a source occupying a spherical space, as shown above. We measure correlations between field points on a measurement sphere surrounding the source. The 3D van Cittert-Zernike theorem for this geometry takes the form $W(\Delta \mathbf{s}, v) = \iiint S(\mathbf{r}, v) e^{j2\pi \frac{\Delta \mathbf{s} \cdot \mathbf{r}}{\lambda}} d^3 \mathbf{r}$, where $W(\Delta \mathbf{s}, v)$ is the cross spectral density between the fields drawn from points separated by Δs on the measurement sphere. $S(\mathbf{r}, v)$ is the 4D source density.

Efficient source state estimation is based on the idea that one wishes to estimate some particular projection of $S(\mathbf{r}, v)$, rather than an isomorphic map. In this case, one may choose to modulate $W(\Delta s, v)$ on measurement. If one multiplies $W(\Delta s, v)$ by a modulating function and integrates one obtains

$$\int W(\Delta \mathbf{s}, v) \psi(\Delta \mathbf{s}, v) d^3 \Delta s \, dv = \int S(\mathbf{r}, v) V(\mathbf{r}, v) d^3 \mathbf{r} \, dv,$$

where $V(\mathbf{r}, v) = \iiint \psi(\Delta \mathbf{s}, v) e^{j2\pi v \frac{\Delta \mathbf{s} \cdot \mathbf{r}}{c}} d^3 \mathbf{r}$, meaning that we obtain a arbitrary projection of the source state by appropriately weighting the mutual coherence measurement.

As part of the 4Is project, we experimentally explored the possibility of efficient abstraction of source primitives by using volume holographic filters and by filtering of nonlocal projections from RSI data. The RSI experiments are discussed in Jason Gallicchio's dissertation in the appendix. Volume holographic sensor systems for confocal prefilters and other efficient sensor systems are described in attached manuscripts.

Optimal filtering schemes for target analysis, as in direct measurement of specific spatio-spectral target features, requires further analysis of the range of holographic, diffractive and sensor plane filtering operations. This approach may be particularly attractive for identification and tracking of distributed biological and chemical targets.

Limits of Multiplex Imaging

Dramatic increases in processing power have made Fourier, Hadamard and other multiplex coding schemes increasingly attractive for spatial sensor systems. Multiplex system analysis common to spectral multiplex sensors is not directly applicable to spatially complex systems, however, because the constant radiance theorem geometrically restricts spatial multiplexing. The constant radiance theorem effectively couples sensor size and multiplexing, meaning that as more modes are multiplexed on measurement sensor size must grow to maintain quantum efficiency.

We have recently published an analysis of the constant radiance theorem in this context in Optics Letters. This analysis is included as an appendix.

Technology Development and Transfer

Technologies developed under 4Is are being transferred to military, security, biomedical and control applications. The basic sensor science developed under 4Is directly lead to applied programs under the DSO, TTO and MTO offices at DARPA aimed at multidimensional microscopies, ground sensor and efficient hyperspectral imaging systems. The integrated sensing and processing community grew dramatically over the three year life of this program and has spawned a new Optical Society of America topical meeting, a week-long symposium on "Frontiers of Imaging" at the 2002 Optical Society Annual meeting and many new research projects around the nation. While relatively few advanced computational imaging systems are currently deployed, events are moving very quickly in this area.

The birefringent sampling field sensor is the only patent currently pending derived from the 4Is program, but several additional disclosures from technologies related to this program have been filed with the Duke University Office of Science and Technology in response to advances under the multidimensional microscopy plan.

As mentioned above, technologies developed under 4Is form the basis of two small companies, Phase Optics, Inc. and Distant Focus Corporation. Phase Optics (www.phaseoptics.com) has licensed SFS technology from the University of Illinois and is building a prototype metrology system for manufacturing process control based on this sensor. Distant Focus (www.distantfocus.com) is working with Raytheon on large sensor arrays and multiplex spatio-spectral tracking systems.

Relationships developed through the 4Is program with the Army Research Lab and with the Air Force Research Lab are continuing, visits by the program PI to ARL in Maryland and AFRL in Ohio are planned for early 2002. Both labs were well represented at the Integrated Computational Imaging Systems OSA topical meeting in the fall of 2002, at which opportunities derived from this program were discussed.

Opportunities

As indicated in the previous sections, 4Is achieved dramatic demonstrations that the information gathering and image formation roles of optical imaging systems could and should be separated. As is often the case, however, this investigation uncovered more unsolved mysteries than it solved. Until very recently, sensor system design was extremely ad hoc. Different sensors were compared only with sensors based on similar design principles. In the next generation, sensor systems of different classes will increasingly be compared based on task-specific functionality.

Throughout this report we have emphasized the term "sensor" over "imaging system." We make this distinction because in most cases the output of the system is not an image,

rather it is a control or alarm signal associated with target analysis. Until recently, sensors were designed by post processing of imaging system data. We hope that the most significant contribution of the 4Is project will be broader realization that it is possible to design physical layer components to sense data specific to sensor tasks and that integrated physical and mathematical design will dramatically improve sensing.

Personnel

Graduate research assistants partially supported under this program included:

Daniel Marks, Remy Tumbar, Jason Gallichio, Prasant Potuluri, A. J. Johnson and Evan Cull.

These students completed 7 master of science dissertations and 2 Ph. D. dissertations using results from this program. These theses are listed below in the publications section.

Staff and postdoctoral fellows supported under this program included:

Ronald Stack (currently president of Distant Focus Corporation), Matt Fetterman George Barbastathis (currently Assistant Professor of Mechanical Engineering at MIT), Rick Morrison (CSO of Distant Focus), and Michal Balberg

Faculty contributing to this program included

David J. Brady, David C. Munson, Jr., Eric Michielssen

Publications

Copies of many of the publications derived from this program are attached in the appendix.

Manuscripts derived in whole or in part from this program remaining in press include:

1. Brady, D. J., *Multiplex sensors and the constant radiance theorem,* to appear in Optics Letters
2. Marks, D. L., R. Stack and D. J. Brady, *Digital refractive distortion correction using the astigmatic coherence sensor,* submitted to Applied Optics
3. Johnson, A. J., D. L. Marks, D. Munson and D. J. Brady, *Surface abstraction from tomography of opaque objects using the CLEAN algorithm,* in preparation
4. Tumbar, R. and D. J. Brady, *Birefringent sampling field sensors,* in preparation.
5. Gallicchio, J., D. J. Brady, E. Cull, *Spatio-spectral triangulation of point sources using the rotational shear interferometer*

13

Published articles derived in whole or in part from this program include:

1. Potuluri, P., M.R. Fetterman, and D.J. Brady, *High depth of field microscopic imaging using an interferometric camera.* Optics Express, 2001. **8**(11): p. 624-630.
2. Marks, D.L., R. Stack, A.J. Johnson, D.J. Brady, and D.C. Munson, *Cone-beam tomography with a digital camera.* Applied Optics, 2001. **40**(11): p. 1795-1805.
3. Tumbar, R., R.A. Stack, and D.J. Brady, *Wave-front sensing with a sampling field sensor.* Applied Optics, 2000. **39**(1): p. 72-84.
4. Marks, D., M. Fetterman, R. Stack, and D.J. Brady. *Spectral tomography from spatial coherence measurements.* in *Proceedings of SPIE - The International Society for Optical Engineering.* 2000. San Jose CA Bellingham WA: Society of Photo-Optical Instrumentation Engineers.
5. Marks, D.M., R.A. Stack, and D.J. Brady, *Astigmatic coherence sensor for digital imaging.* Optics Letters, 2000. **25**(23): p. 1726-1728.
6. Fetterman, M.R., E. Tan, L. Ying, R.A. Stack, D.L. Marks, S. Feller, E. Cull, J.M. Sullivan, D.C. Munson, S.T. Thoroddsen, and D.J. Brady, *Tomographic imaging of foam.* Optics Express, 2000. **7**(5): p. 186-197.
7. Balberg, M., G. Barbastathis, S. Fantini, and D.J. Brady. *Confocal imaging through scattering media with a volume holographic filter.* in *Proceedings of SPIE - The International Society for Optical Engineering.* 2000. San Jose CA Bellingham WA: Society of Photo-Optical Instrumentation Engineers.
8. Marks, D.L., R.A. Stack, D.J. Brady, and J. van der Gracht, *Three-dimensional tomography using a cubic-phase plate extended depth-of-field system.* Optics Letters, 1999. **24**(4): p. 253-255.
9. Marks, D.L., R.A. Stack, and D.J. Brady, *Three-dimensional coherence imaging in the Fresnel domain.* Applied Optics, 1999. **38**(8): p. 1332-1342.
10. Marks, D.L., R.A. Stack, D.J. Brady, D.C. Munson, and R.B. Brady, *Visible cone-beam tomography with a lensless interferometric camera.* Science, 1999. **284**(5423): p. 2164-2166.
11. Barbastathis, G. and D.J. Brady, *Volume holographic imaging of three-dimensional objects.* Proceedings of SPIE - The International Society for Optical Engineering, 1999. **3633**: p. 170-181.
12. Barbastathis, G. and D.J. Brady, *Spatio-spectral tomography of luminescent objects with volume holograms.* Proceedings of SPIE - The International Society for Optical Engineering, 1999. **3749**: p. 398-399.
13. Barbastathis, G., M. Balberg, and D.J. Brady, *Confocal microscopy with a volume holographic filter.* Optics Letters, 1999. **24**(12): p. 811-813.
14. Barbastathis, G. and D.J. Brady, *Multidimensional tomographic imaging using volume holography.* Proceedings of the Ieee, 1999. **87**(12): p. 2098-2120.
15. Marks, D.L. and D.J. Brady, *Three-dimensional source reconstruction with a scanned pinhole camera.* Optics Letters, 1998. **23**(11): p. 820-822.
16. Wu, Y. and Munson D.C, Jr. *Wide-angle ISAR passive imaging using Smoothed Pseudo Wigner-Ville distribution.* in *IEEE National Radar Conference - Proceedings.* 2001. Atlanta, GA.

17. Xiao, S., Munson D.C, Jr., S. Basu, and Y. Bresler. *An N^2 log N back-projection algorithm for SAR image formation.* in *Conference Record of the Asilomar Conference on Signals, Systems and Computers.* 2000. Pacific Grove, CA.

M.S. Theses

1. Tumbar, R., *Field sampling and shearing wavefront sensors*, in *Electrical and Computer Engineering.* 1998, University of Illinois. p. v, 74 leaves, bound.
2. Marks, D.L., *Fresnel zone three-dimensional coherence imaging*, in *Electrical and Computer Engineering.* 1998, University of Illinois. p. viii, 77 leaves, bound.
3. Guo, J., *Holographic and Polarization Methods for Optical Field Analysis*, in *Electrical and Computer Engineering.* 1998, University of Illinois: Urbana. p. 116.
4. Johnson, A.J., *Patch response of cone-beam tomography*, in *Electrical and Computer Engineering.* 1999, University of Illinios at Urbana-Champaign: Urbana. p. v, 40 leaves, bound.
5. Gallicchio, J.R., *Spatio-spectral triangulation of visible and infrared point sources using a portable rotational shear interferometer*, in *Electrical and Computer Engineering.* 2001, University of Illinois.
6. Zheng, Y., *Information theoretic design and optimization of imaging systems*, in *Electrical and Computer Engineering.* 2001, University of Illinois.
7. Potuluri, P., *High depth of field and aberration control in microscopy with coherence imaging systems*, in *Electrical and Computer Engineering.* 2001, University of Illinois: Urbana. p. 43.

Ph. D. Theses

1. Marks, D.L., *Four-dimensional coherence sensing*, in *Electrical and Computer Engineering.* 2001, University of Illinois: Urbana. p. 127.
2. Tumbar, R., *High spatial bandwidth wavefront sensing by sampling and spatial multiplexing*, in *Electrical and Computer Engineering.* 2001, University of Illinois: Urbana.

Appendix: Publication Reprints

Publications on Sampling Field Sensors

Sections from High Spatial Frequency Wavefront Sensing by Sampling and Spatial Multiplexing

Thesis submitted in partial fulfillment of the requirements for the Ph. D. degree

By Remy Tumbar

CHAPTER 1. INTRODUCTION AND OVERVIEW

1.1   Introduction

Optical sensors are used in a broad range of applications. These include classical imaging systems, tomographic systems, and interferometric and noninterferometric sensors for amplitude, phase, or polarization. The imaging paradigm itself has changed from producing a point-to-point mapping of the intensity distribution of an object to producing an intensity representation which is not necessarily a point-to-point map. Recent developments in computational hardware and software have allowed a wider use of computational imaging systems and sensors, which rely in extensive postdetection processing of the sensor data in order to reconstruct the object or the sensed parameter distribution. Examples of such imaging systems are wavefront sensors like the Shack-Hartmann sensor (SHS) [1] and the sampling field sensor (SFS) [2], tomographic systems, and interferometric imaging systems like the rotational shearing interferometer (RSI) [3]. The use of computers to extensively process sensor data allows greater freedom in the sensor design process resulting in a better accomplishment of the given imaging task. Ideally one would optimize the optical sensing system considering the object or field parameter to be "imaged," system resources, input/output field statistics, and noise sources. This would result in an optical sensor tailor-made for each particular situation. The classical imaging goal has thus shifted towards efficiently and reliably getting the maximum amount of information from the input optical field.

As a particular case of optical sensors we consider wavefront sensors, which measure the phase and amplitude of the input optical field. Their traditional use, for measuring the quality of optical lenses, has diversified in light of the above-mentioned computational imaging trend. They are now used for imaging through turbulent media using adaptive optics

[4] or deconvolution from wavefront sensing [5], or for tomographic reconstructions of transparent objects [6]. Other applications include optical testing, eye aberration measurements [7] for eye surgery, and diffractive element characterization [8]. Wavefront sensors have long posed problems in their use for two reasons. First the phase of the optical field is not available directly, but it has to be coded into intensity measurements. In some cases this leads to multiple-shot interferometric sensors, which cannot be used in real time. Second, most of them, especially interferometric sensors, are very sensitive to mechanical vibrations encountered in a nonlaboratory environment. The latter has prevented their widespread use as well as their consideration for even more applications. The SHS overcomes both of these problems. This fact has driven its proliferation in various applications such as astronomic imaging, laser corrective eye surgery [7], laser beam and lens quality measurements, and others. However, the SHS has a low information capacity when measured as the space-bandwidth product [2]. We propose a new type wavefront sensor, the SFS. The SFS is compact and vibration-insensitive. It has a high space-bandwidth product and it takes all the data in one shot, making it a strong candidate for real-time applications. Its principle of operation and a description of its first-generation implementation are described in a recent paper [2].

1.2 Sampling and Spatial Multiplexing

Interferometric shearing wave-front sensors use different methods to create two laterally shifted or sheared copies of the input, which they interfere at the output plane. In addition, the relative phase/optical path delay of one of the copies with respect to the other is changed and multiple phase-shifted frames/shots of the output are recorded. This is done by moving mirrors, moving gratings, moving wave plates, and other means. These operations may remove degeneracies (like in the estimating the phase angle from its cosine value, which is measured by interferometers) or they may increase the accuracy of the estimate in the presence of noise. The so-called phase-shift algorithms are used for that purpose. Usually this is done for each lateral shear direction separately.

The wavefront map is reconstructed from measurements of the phase differences between adjacent points in the input aperture separated by the shearing distance in the direction of the shear. A problem arises when the input changes (e.g., a pulsed or variable field) or when the system also changes uncontrollably between the measurements, thus reducing the accuracy of the result. A common solution used to achieve one-shot detection in multishot systems is the use of multipath system implementations. This amounts to splitting the input field and sending it through different optical paths into different subsystems with specific functions, such as shear along x, or shear along y, or specific phase delays [9].

We use sampling and space multiplexing to generate the required phase shift and shear diversity information in fewer frames than usually necessary, even in a single frame. In addition, the shear and phase-shift diversity information is generated along an essentially common optical path. Thus, the present method eliminates the disadvantages related to multipath methods, which are the current solution to turning multishot systems into single-shot ones. These disadvantages include increased sensitivity to vibrations, increased complexity and size of the system, and poorer reliability (due to increased complexity). The wavefront sensing systems we describe in this work are compact and much less sensitive to vibrations and misalignment than other interferometric sensors

By sampling the input field we obtain a sparse version with blocked regions separating the samples. We then use the empty space between the samples to generate additional information about the input such as shearing interferometric measurements on multiple directions and phase-shifted measurements. We call diversity information the set of multiple measurements used in estimating the input. This work presents two different ways of generating the diversity information.

The sampled field diffracts, through free-space, creating overlapping patterns at the output of the device. The patterns produce sheared interferometric measurements along the Cartesian direction and their phase curvature gives additional phase-diversity information. This implementation of our method suffers from very low light throughput, small angular

bandwidth, and larger cross-talk between the measurements than the second method. It is, however, extremely cheap, since it can be done by mounting a sampling mask about a millimeter away in front a CCD chip.

The second implementation generates the diversity information by imaging the input or sampling plane to the output of the system rather than letting it diffract in free space. Multiple, appropriately modified copies of the samples are produced at the output by using birefringent plates that split the optical field in multiple components propagating in different directions. The geometrical layout of the patterns allows overlap, which further produces multiple phase-shifted and sheared interferometric measurements. Due to imaging the sampled field, throughput two orders of magnitude larger than that of the previous method is achieved while minimizing the cross-talk. The new system achieves phase sensitivity comparable to high-performance phase shift systems while being virtually insensitive to vibrations.

1.3 Overview of Previous Work on Wavefront Sensors and Phase-Estimation and Reconstruction Algorithms

Wavefront sensors include interferometric and noninterferometric systems [10]. Shearing interferometers, point diffraction interferometers [11], and the pseudo-phase-conjugate interferometer [12] are a few examples of interferometric wavefront sensors. Non-interferometric wavefront sensors include the Shack-Hartmann sensor (SHS) [1] and the curvature sensor [13] with the SHS being the most commonly used wavefront sensor. The SHS is constructed by mounting an array of lenses in front of an array of detectors. The detectors determine the position of the focal spot intensity centroid of each lens. The set of centroid positions for a normally incident plane wave serves as a zero reference. The offsets between the detected centroids for an arbitrary input wavefront and the zero reference positions provide measures of the average wavefront tilt coefficients over each lens sub-aperture. The input wavefront is then reconstructed via data reduction techniques using these average tilt measurements.

Optical wavefront sensors code phase information into intensity information in their output, making two additional steps necessary compared to lower frequency phase sensors. The first step is the phase estimation: use the intensity measurements in the sensor output to reliably and efficiently estimate the input phase or an invertible representation of the input phase. Usually, this step gives the input phase in the form of exact finite differences between adjacent points on a Cartesian grid for shearing interferometers and the SFS. In the SHS case the average tilt coefficients given by the sensor are estimated using the intensity centroids in the output plane of the sensor and not regular phase estimation algorithms. The second step is the phase reconstruction. This amounts to reconstructing the input phase distribution as a continuous function from the results at the previous step. Since both steps apply in the SFS case, we consider them to be connected to the problem of SFS design and testing.

The most important class of phase estimation algorithms in optics is that of phase-shifting interferometry algorithms. Creath [14] gives a good review of the subject. The idea is to sample the waveform

$$I(\delta) = I_b + I_0 \cos(\varphi + \delta) \tag{1.1}$$

for different values of δ and use the samples to estimate φ in the $[-\pi, \pi)$ interval. Given that I_b, I_0, and φ are the unknowns, one needs only three measurements to estimate these unknowns with infinite accuracy if there are no sources of noise or systematic errors. However, there are significant systematic as well as random errors in the actual physical measurements. The main trend in phase-shifting algorithm design has been to design algorithms which are insensitive to a given set of systematic errors, at the expense of having more than three samples of Eq. (1.1). In general, the phase estimate is obtained from multiple measurements using

$$\varphi = \tan^{-1} \left(\frac{\sum_{k=1}^{M} b_k I(\delta_k)}{\sum_{k=1}^{M} a_k I(\delta_k)} \right) \tag{1.2}$$

where δ_k are the phase shifts, a_k and b_k the coefficients, and $I(\delta_k)$ the measurements in an M-frame algorithm. Waveform nonlinearity and phase-shift miscalibration are the main sources

of errors considered. The emphasis given to systematic errors stemmed from the poor repeatability of the phase-shift devices and from the nonlinearity of focal plane detectors [14]. One can eliminate the first problem by using separate motion detectors, either in the form of a parallel interferometer or a magnetic position sensor, for example. Spatially nonuniform phase shifts are harder to compensate. Some phase estimation algorithms [15] eliminate systematic errors up to a given order by considering the Fourier expansion of $I(\delta)$ in $\exp(im\varphi)$ and making the coefficients for $m \neq 1$ equal zero. This causes the roots of a certain polynomial to be on the unit circle.

Another approach [16] is to add error terms to the ideal form of Eq. (1.1) either in the argument of the cosine (phase calibration errors or nonlinear phase shifter) or as nonlinearity of the I_0 term due to multiple beam interference or nonlinearity of the intensity detector. By forcing the coefficients of the error terms to be zero one obtains a set of equations for the coefficients of the phase estimating algorithm. Hibino [17] has proven that algorithms that compensate for both the harmonics of the nonlinearity and the phase shift miscalibration are, in general, more susceptible to errors caused by random noise. The solution he suggested was to use even more measurements than the minimum number required by the elimination of systematic errors. This would increase the signal-to-noise ratio (SNR) but not completely eliminate the errors.

A third approach is to consider all the errors to be random and to use maximum-likelihood estimation theory to estimate the phase from a given set of measurements [18], [19]. Rogala [18] also used the Cramer-Rao lower bound as a performance measure of the experimental setup used to identify the best way of taking the measurements. This last approach has found only a few followers for two reasons, in our opinion. First, the results of the estimation can depend rather strongly on the noise model used. Using the wrong noise model can give poorer results than expected. In particular, Rogala's choice of the noise models was rather ad-hoc. Secondly, the statistics of systematic errors may be very difficult to estimate due to their large high-order moments. This hinders the proper use of statistical estimation methods like the ML method. A stronger (exact) cancellation of errors is

required, which is provided by the deterministic algorithms obtained by the first and second design methods.

Phase reconstruction from sensor data appears in two forms. The first is the phase map reconstruction from exact or average finite differences that appears mostly in shearing interferometry problems. Since the SFS outputs data similar to a shearing interferometer, this problem is connected to this work. The problem of unwrapping two-dimensional phase maps is related to reconstruction from phase differences because the usual approach is first to form the phase differences and then to reconstruct the unwrapped phase map from these differences. Therefore, much of the work done on phase unwrapping algorithms applies to our problem of phase reconstruction from finite differences. The second form of phase reconstruction from sensor data can be viewed as an inverse problem: one uses the intensity distributions in the output plane of a sensor to reconstruct the input phase map. It is not directly related to the SFS, at least in its implementations considered for this work. It is, however, an important problem in the general setup of wavefront sensor design. Its study will point out to advantages and disadvantages of using certain types of wavefront sensors and wavefront sensing methods in certain situations. As a result, one can design improved wavefront sensors and wavefront reconstruction algorithms.

Historically, there are two classes of wavefront reconstructions from phase differences: zonal [20, 21] and modal reconstructions [22-24]. It is easy to show, however, that the classical zonal reconstruction method is in fact a modal reconstruction when the expansion modes are *sinc* functions and the number of expansion modes is equal to the number of measurement points within the sensor's aperture. Equations (1.3) and (1.4) show a matrix of zonal and modal representations, respectively, of the wavefront reconstruction problem:

$$\mathbf{g} = \mathbf{A}\phi_d \tag{1.3}$$

$$\mathbf{g} = \mathbf{A}\mathbf{Z}\mathbf{c} \tag{1.4}$$

where ϕ_d is the grid of sampled wave frontphase values, **c** is the set of coefficients representing the wave front phase in the chosen basis, **Z** is the operator having the basis vectors as columns, and **A** is the finite difference operator, given by

$$
\mathbf{A} = \begin{pmatrix}
1 & -1 & & & & & & & \\
 & 1 & -1 & & & & & & \\
 & & & 1 & -1 & & & & \\
 & & & & 1 & -1 & & & \\
 & & & & & & 1 & -1 & \\
 & & & & & & & 1 & -1 \\
1 & & & -1 & & & & & \\
 & 1 & & & -1 & & & & \\
 & & 1 & & & -1 & & & \\
 & & & 1 & & & -1 & & \\
 & & & & 1 & & & -1 & \\
 & & & & & 1 & & & -1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{pmatrix}
\tag{1.5}
$$

for the case of a 3×3 phase map. A zonal reconstruction will estimate the vector ϕ_d of sampled phase values, whereas a modal one will find the coefficient vector **c**. In general, the size of the modal coefficient vector is smaller than the size of the phase array. However, a choice of the modal basis in the form of *sinc* functions, with the size equal to the number of points in the array, makes Eqs. (1.3) and (1.4) the same. The purpose of using modal representations is only partially revealed by the problem of wavefront reconstruction from the phase differences. The remaining part will be given when discussing the problem of wavefront reconstructions from sensor data. For now the motivation relies on the fact that modal representations may use lower-dimensionality spaces to estimate the input. This assumes that the input has in fact fewer degrees of freedom than its number of grid points. If one adds noise in the left-hand side (LHS) of Eqs. (1.3) and (1.4), the reconstruction error will not increase with more grid points as in the case of zonal reconstructions where it scales as $1/\pi\ln(N)$. The scaling law has been explained by Noll [25] by using Green functions and by Menikkoff [26] by using a cosine series representation of the finite difference operator. In short, both explanations boil down to the $1/k^2$ dependence of the spatial power density of the

wavefront, $<\phi_d^*(\mathbf{k})\phi_d(\mathbf{k})>$. White noise in the phase differences will produce a $1/k^2$ dependence of the error. Increasing the number of points gives rise to smaller wave numbers and therefore to larger errors. If the input that had given rise to the phase difference data is not in the range space of the $\mathbf{Z}$ operator, the reconstruction will be biased (increased modeling error). Two other problems with modal reconstructions are aberration coupling and aberration aliasing. The nonorthogonality of the expansion modes causes the first problem, while their linear dependence causes the second [23]. Both are effects of the finite sampling, which makes higher-order modes, having more structure, similar to lower-order ones. The basis most widely used is that of Zernike polynomials because it represents well the aberration terms in an optical system. It has been argued that it is not a good basis for more general wavefront reconstruction problems [27]. In wavefront sensing of Kolmogorov turbulence (atmospheric adaptive optics), for example, the best basis is the Karhunen-Loeve set. This approach can be used in a more general situation given that the statistics of the input wavefront are known. Legendre polynomials and Fourier sets are also widely used due to their orthogonality on square domains.

Both the zonal and the modal problem are usually solved in a least-squares sense. This is common to the phase unwrapping techniques that first form the phase differences and then do a minimization of the least-squares error. The advantage of this technique is in allowing the use of fast elliptic solvers [28]. It has been argued (mainly in phase unwrapping works) that the L^2-norm error minimization produces biased reconstructions. This can be explained in two ways. One is to note that the phase reconstruction problem becomes that of solving Poisson's equation with an appropriately constructed source term and Neumann boundary conditions [28]. The source term is large in regions where there are large phase differences such as noisy or aliased (phase difference larger than π) regions. Solving the Poisson's equation is a deconvolution problem. The deconvolution kernel is a low-pass filter since the Laplacean operator is a high-pass one. This explains the long range of the distortions in the reconstruction. A more interesting explanation has been suggested by Fried [29] using the concept of branch points. Noise in the phase estimation process and aliasing in the actual input phase map give rise to a solenoidal component of the phase field. This

does not appear in the source term of the corresponding Poisson equation constructed in the least-squares approach. It can be added by postulating its existence and adding a dipole term to the reconstructed phase map [29], [30]. It has also been argued that using more general L^P norms, especially L^0 and L^1, will give better reconstructions by forcing the gradient of the reconstructed map to follow that of the input. The use of the L^P norms in the context of the solenoidal phase term has not been studied.

Modal reconstructions are also used in the context of a more general problem: reconstructing the input phase map from intensity distributions in the output plane of the wavefront sensor. The SHS gives the best example because of its wide practical use. It is constructed by mounting an array of lenses in front of an array of detectors, like a CCD camera. Each lens focuses the wavefront incident on it onto a corresponding region of the detector array. From elementary diffraction theory one can show that the field at the detector plane is a scaled Fourier transform of the field incident onto the respective lens [31]. Therefore, the SHS output field consists of a set of windowed Fourier transforms of the input field. The actual sensor data is an absolute value squared of the output field distribution. The centroid of the intensity distribution in each lens output region of the output plane is proportional to the average tilt coefficient over that of the input wavefront over the lens sub-aperture [1]. The input phase map is reconstructed by either assuming that the average tilts are good estimates of the actual discrete phase differences or by considering a modal representation of the input, computing the range space of its corresponding sensor output, and projecting the average tilt measurements onto that range space. We have a zonal reconstruction in the first case and a modal one in the second. In this case, the advantage of the modal reconstruction consists in providing an extended representation of the input wavefront, beyond the combination of subaperture tilts. This is in contrast to the problem of modal phase reconstruction from the phase differences where a better representation of the input amounted to using a smaller basis set, thereby limiting the influence of noise. Also, the representation of the input by using *sinc* functions does not make modal reconstruction equivalent to the zodal one in the way it does for reconstruction from phase differences. The reason is the nonlinear dependence between the input and the output measurements and the fact that only the subaperture averaged wavefront tilts are estimated from the measurements. A more radical approach to this inverse problem is to use all the information in the output

plane of the sensor, not only the intensity centroids [32]. This results in recovering more of the structure of the input wavefront. This approach is, however, very demanding from a computational standpoint. Given the set of applications where the SHS is the best choice, a refinement of this technique is needed.

1.4 High-Space-Bandwidth Wavefront Sensing by Sampling and Spatial Multiplexing

The space-bandwidth product (SBP) of an optical system can be defined in different ways [33]. One is to consider the space domain, another one is to consider the spatial frequency domain, and the third is to consider the space/spatial frequency domain by using the Wigner transform. All the above produce a number which is invariant under the transforms related to imaging with a lens, free-space diffraction, and under the Fourier transform [33]. Therefore this number is a good measure of the information capacity of an imaging system. The space domain definition of the SBP is

$$SBP = 2B_x B_y L_x L_y + 2 \cong 2B_x B_y L_x L_y \tag{1.6}$$

where the factor of 2 applies to coherent sensors only and the last term is the minimum number of degrees of freedom required to specify a coherent field. The large number of degrees of freedom in optical systems justifies the approximation in Eq. (1.6). Equation (1.6) does not account for the state of polarization. The factors B_x and B_y are the spatial bandwidths in the x and y directions, respectively. The factors L_x and L_y are the dimensions of the sensor aperture in the x and y directions, respectively. The spatial bandwidths are less or equal to the Nyquist limits given by sampling in both Cartesian directions, as the sensor may have additional bandwidth limitations. Consider the Nyquist limits to be $1/2h_x$ and $1/2h_y$ and the actual spatial bandwidths to be the fractions $g_x/2h_x$ and $g_y/2h_y$, with g_x and g_y less than 1. Then the SBP of a given sensor will be

$$SBP = g_x g_y \frac{L_x L_y}{2h_x h_y} \tag{1.7}$$

As explained earlier in this chapter, the most common examples of optical wavefront sensors are phase-shift shearing interferometers and Shack-Hartmann sensors. We

performed numerical simulations on the SHS (see Chapter 2) and found that $g_x = g_y \approx 0.1$. This is due mainly to the fact that the SHS outputs average phase differences for each sampling period. The mean value theorem can be used to prove that the averages come closer to the actual differences as the bandwidth of the signal decreases. Ten times oversampling gave about 20% reconstruction error. Also, the lenslets in the lenslet array of the SHS are difficult to make with a lateral dimension less than approximately 50 μm, which gives the sampling distances h_x and h_y for this case. We therefore have that

$$SBP(\text{SHS}) = 0.01 \frac{L_x L_y}{5000} \tag{1.8}$$

where L_x and L_y are now in microns.

Interferometers detect fringe patterns as described by Eq. (1.1). Here the oversampling is required by the smoothing of the measured data through integrating the fringes by the finite-sized camera pixels. This is somewhat similar to the case of the SHS, so we will use a rough estimate of 10 times oversampling per spatial direction. On the other hand, the sampling distance in a phase-shift interferometer can be as small as the pixel size on the CCD used to get the interferometric data, which can be as small as 5 μm. We therefore have that

$$SBP(\text{interferometer}) = 0.01 \frac{L_x L_y}{50} \tag{1.9}$$

Equations (1.8) and (1.9) show that interferometers have about one hundred times the SBP of Shack-Hartmann sensors.

The new type of interferometric sensor that we describe in this work, the SFS, outputs data similar to a phase-shift shearing interferometer. The first implementation, which uses free-space diffraction fan-out, has an additional limitation on the sampling hole size which limits the sampling distance to about 100 μm per sampling direction (see Chapter 2). Given that this limit is satisfied, the diffractive SFS can detect fields with bandwidths up to the Nyquist limit, which means it has an SBP of

$$SBP(\text{diffractive SFS}) = \frac{L_x L_y}{20000} \tag{1.10}$$

which is about 25 times larger than the SBP(SHS). The second SFS implementation, which uses birefringent fan-out, does not have the additional limitation and can have an SBP equal to that of regular interferometers, thus about 100 times better than the SBP(SHS). None of the above SBP figures considers the effect of noise and the finite dynamic range of CCD systems, so they are only rough estimates used for the purpose of comparing the different sensor technologies. More research is required to obtain accurate numbers for the relative improvements of the information capacity of one technology with respect to another.

1.5   Overview of This Work

We study two different implementations of a new type of wavefront sensor, the SFS. Chapter 2 describes the SFS with fan-out through free-space diffraction. We present the concept, the design procedure, numerical simulations of the system, and experimental results that validate the concept. Chapter 3 describes SFS with fan-out through imaging the sampled field through a set of birefringent crystals. We present the concept and a theoretical calculation of the design parameters as well as the design procedure and specific examples of fan-out generation. Chapter 4 gives a detailed account of the experimental verification of the concepts described in Chapter 3. We conclude in Chapter 5 with a discussion of the results of this work.

1.6   References

[1]   R. K. Tyson, *Principles of Adaptive Optics*. Boston: Academic Press, 1991.

[2]   R. Tumbar, R. A. Stack, and D. J. Brady, "Wave-front sensing with a sampling field sensor," *Applied Optics*, vol. 39, pp. 72-84, 2000.

[3]   D. L. Marks, R. A. Stack, and D. J. Brady, "Three-dimensional coherence imaging in the Fresnel domain," *Applied Optics*, vol. 38, pp. 1332-1342, 1999.

[4]   M. C. Roggemann, B. M. Welsh, and R. Q. Fugate, "Improving the resolution of ground-based telescopes," *Reviews of Modern Physics*, vol. 69, pp. 437-505, 1997.

[5]   B. M. Welsh and M. C. Roggemann, "Signal-to-noise comparison of deconvolution from wave-front sensing with traditional linear and speckle image reconstruction," *Applied Optics*, vol. 34, pp. 2111-2119, 1995.

[6] M. C. Roggemann, B. M. Welsh, P. J. Gardner, R. L. Johnson, and B. L. Pedersen, "Sensing three-dimensional index-of-refraction variations by means of optical wavefront sensor measurements and tomographic reconstruction," *Optical Engineering*, vol. 34, pp. 1374-1384, 1995.

[7] S. A. Klein, "Optimal corneal ablation for eyes with arbitrary Hartmann-Shack aberrations," *Journal of the Optical Society of America A-Optics & Image Science*, vol. 15, pp. 2580-2588, 1998.

[8] M. Zajac and B. Dubik, "Measurement of wavefront aberrations of diffractive imaging elements," *SPIE Int. Soc. Opt. Eng. Proceedings of Spie the International Society for Optical Engineering*, vol. 3320, pp. 237-241, 1998.

[9] A. L. Weijers, H. van Brug, and H. J. Frankena, "Polarization phase stepping with a Savart element," *Applied Optics*, vol. 37, pp. 5150-5155, 1998.

[10] J. M. Geary, *Introduction to Wavefront Sensors*. Bellingham, Washington, SPIE Optical Engineering Press, 1995.

[11] R. N. Smartt and W. H. Steel, "Theory and application of point-diffraction interferometers (telescope testing)," *Japanese Journal of Applied Physics*, vol. 14, pp. 351-356, 1975.

[12] Y. Baharav, B. Spektor, J. Shamir, D. G. Crowe, W. Rhodes, and R. Stroud, "Wavefront sensing by pseudo-phase-conjugate interferometry," *Applied Optics*, vol. 34, pp. 108-113, 1995.

[13] F. Roddier, "Curvature sensing and compensation: a new concept in adaptive optics," *Applied Optics*, vol. 27, pp. 1223-1225, 1988.

[14] K. Creath, "Phase-measurement interferometry techniques," in *Progress in Optics. Vol. XXVI*, pp. 349-393, 1988.

[15] Y. Surrel, "Design of algorithms for phase measurements by the use of phase stepping," *Applied Optics*, vol. 35, pp. 51-60, 1996.

[16] K. Hibino, B. F. Oreb, D. I. Farrant, and K. G. Larkin, "Phase-shifting algorithms for nonlinear and spatially nonuniform phase shifts," *Journal of the Optical Society of America A-Optics & Image Science*, vol. 14, pp. 918-930, 1997.

[17] K. Hibino, K. G. Larkin, B. F. Oreb, and D. I. Farrant, "Phase-shifting algorithms for nonlinear and spatially nonuniform phase shifts: reply to comment," *Journal of the Optical Society of America A-Optics & Image Science*, vol. 15, pp. 1234-1235, 1998.

[18] E. W. Rogala and H. H. Barrett, "Phase-shifting interferometry and maximum-likelihood estimation theory," *Applied Optics*, vol. 36, pp. 8871-8876, 1997.

[19] E. W. Rogala and H. H. Barrett, "Phase-shifting interferometry and maximum-likelihood estimation theory. II. A generalized solution," *Applied Optics*, vol. 37, pp. 7253-7258, 1998.

[20] D. L. Fried, "Least-square fitting a wave-front distortion estimate to an array of phase-difference measurements," *Journal of the Optical Society of America*, vol. 67, pp. 370-375, 1977.

[21] R. H. Hudgin, "Wave-front reconstruction for compensated imaging," *Journal of the Optical Society of America*, vol. 67, pp. 375-378, 1977.

[22] R. Cubalchini, "Modal wave-front estimation from phase derivative measurements," *Journal of the Optical Society of America*, vol. 69, pp. 972-977, 1979.

[23] J. Herrmann, "Cross coupling and aliasing in modal wave-front estimation," *Journal of the Optical Society of America*, vol. 71, pp. 989-992, 1981.

[24] W. H. Southwell, "Wave-front estimation from wave-front slope measurements," *Journal of the Optical Society of America*, vol. 70, pp. 998-1006, 1980.

[25] R. J. Noll, "Phase estimates from slope-type wave-front sensors," *Journal of the Optical Society of America*, vol. 68, pp. 139-140, 1978.

[26] A. Menikoff, "Wave-front reconstruction with a square aperture," *Journal of the Optical Society of America A-Optics & Image Science*, vol. 6, pp. 1027-1030, 1989.

[27] R. G. Lane and M. Tallon, "Wave-front reconstruction using a Shack-Hartmann sensor," *Applied Optics*, vol. 31, pp. 6902-6908, 1992.

[28] D. C. Ghiglia and L. A. Romero, "Robust two-dimensional weighted and unweighted phase unwrapping that uses fast transforms and iterative methods," *Journal of the Optical Society of America A-Optics & Image Science*, vol. 11, pp. 107-117, 1994.

[29] D. L. Fried, "Branch point problem in adaptive optics," *Journal of the Optical Society of America A-Optics & Image Science*, vol. 15, pp. 2759-2768, 1998.

[30] W. W. Arrasmith, "Branch-point-tolerant least-squares phase reconstructor," *Journal of the Optical Society of America A-Optics & Image Science*, vol. 16, pp. 1864-1872, 1999.

[31] J. W. Goodman, *Introduction to Fourier Optics*, 2nd edition. New York: McGraw-Hill, 1996.

[32] R. C. Cannon, "Global wave-front reconstruction using Shack-Hartmann sensors," *Journal of the Optical Society of America A-Optics & Image Science*, vol. 12, pp. 2031-2039, 1995.

[33]   A. Lohmann, R. G. Dorsch, D. Mendlovic, Z. Zalevsky, and C. Ferreira, "Space-bandwidth product of optical signals and systems," *Journal of the Optical Society of America A-Optics & Image Science*, vol. 13, pp. 470-473, 1996.

5. DISCUSSION

We have described a new type of wavefront sensor, the SFS. It is, to our knowledge, the first high-accuracy, high-resolution (high SBP), vibration-insensitive, one-shot wavefront sensor. It achieves high accuracy by using a phase-shift shearing interferometric approach in detecting wavefront phase differences. It achieves high resolution by sampling the input wavefront and fanning out the samples to the output plane (space multiplexing). Not only does this provide one-shot detection, but it does it in a common path setup, i.e., the optical paths of the field in different parts of the fan-out pattern go through essentially the same optics. This gives a very compact and vibration insensitive system.

We have designed and tested two implementations of the SFS, one doing space multiplexing (fan-out) through free-space diffraction and the other through imaging through birefringent crystals. In this chapter we summarize and discuss the experimental results, and we compare the two implementations based on their fundamental performance parameters such as signal-to-noise ratio and space-bandwidth product.

Summary and Discussion of Results

In this section we summarize the results of numerical simulations and experiments and comment on their relevance.

Diffractive SFS

We tested the consistency of the SFS method both numerically and experimentally. Numerical simulations proved the wavefront sensing principle of this device. We compared the phase estimation accuracy of the system with that of a Shack-Hartmann sensor, through numerical simulations. Although its low light throughput would give this SFS implementation an SNR 10 times worse than that of an SHS, this would not be a factor in

testing high intensity laser sources, for example. We chose to test the accuracy without factoring in any measurement noise. The numerical simulations showed that the input phase differences could be estimated with 20% accuracy up to the Nyquist limit given by sampling. This is in contrast to the SHS, which achieved 20% accuracy only at about a third of the Nyquist limit (see Table 2.1).

Experimentally, we showed that the signal detected in the phase pixels is proportional to the cosine of the phase difference between the two corresponding adjacent sampling points. However, we were not able to calibrate the system in the experiments mainly because of low SNR, low dynamic range, and low system bandwidth. These prevented us from using input wavefronts with enough diversity to allow the calibration of the system.

The advantages of this implementation are its simplicity and direct application to testing pulsed lasers. Also, we showed that a slightly different design, i.e., placing a tapered-hole mask in front of a CCD array, will result in a very compact, cheap, and vibration-insensitive wavefront sensor. The disadvantages of this implementation are its low light efficiency and small input bandwidth compared to other interferometric setups and to the second SFS implementation.

Birefringent SFS

The main limitations of the previous SFS implementation (low throughput, dynamic range and system bandwidth) are due to free-space diffraction being an isotropic phenomenon when the generation of SFS characteristic patterns requires an anisotropic process. However, similar patterns, although not with all the required properties, can be generated using array generation techniques such as imaging through birefringent plates [1], a highly anisotropic transformation. The idea is to modify the PSF of the imaging system from just one bandlimited *jinc* function to multiple copies of it with additional phase differences between these copies. To our knowledge, previously reported work in array generation using birefringent elements did not consider the control of the phase difference between the fan-out copies of the input.

We showed that a characteristic pattern satisfying the requirements of the SFS can be generated using birefringent elements. We derived the design equations with both the ray optics and the Fourier optics approach. The latter revealed a very important degree of freedom in designing the system. It consisted of using a cascaded imaging system and placing the set of birefringent plates between the second principal plane of the first imaging element and the first principal plane of the second imaging element. This allowed us to magnify only the sampled field in the first step and both the already magnified sampled field and the pattern generated by the birefringent plates in the second imaging step. As a result, we were able to match the geometry of the characteristic pattern generated by a given set of plates with multiple sampling and receiving geometries.

Experimentally, we proved the validity of our design equations. Our tests consisted of measuring the chirped wavefront produced by focusing a well-collimated laser beam with a high-performance laser aplanat. The most important experimental result was proving that the system outputs complete shearing phase-shift interferometric data, i.e., shearing interferometric data along two noncollinear directions with multiple different phase shifts for each shearing direction.

Different shearing interferometric data with low cross-talk was shown in Fig. 4.4 on page 76. The low cross-talk is obvious from the quality of the interferometric fringes. To further prove it, we also measured wavefronts that were chirped only along one direction, which we obtained by focusing a well-collimated laser beam with cylindrical lenses. Figures 5.1 and 5.2 show the results for two different cylindrical lenses. The lenses were placed against the sampling mask so the curvature measured by the system was approximately equal to the focal length of each lens. The measurements showed, as in the case of spherical lenses, accurate estimation of the curvature of the input wavefront. In addition, the detected phase curvature in the direction perpendicular to chirp was zero, which means that there was little cross-talk between the output pixels measuring shearing interferometric data along the two directions, perpendicular and parallel to the direction of the chirp.

Modulation in the direction of the chirp for a 150 mm cylindrical lens

Peaks correspond to F = 149.1 mm

Modulation in the direction perpendicular to chirp for a 150 mm cylindrical lens

Peak corresponds to F = ∞
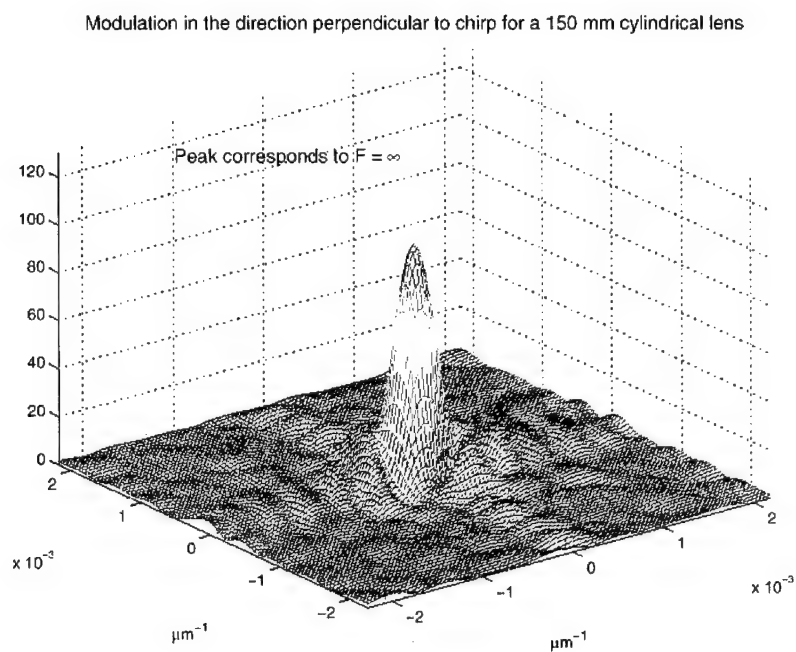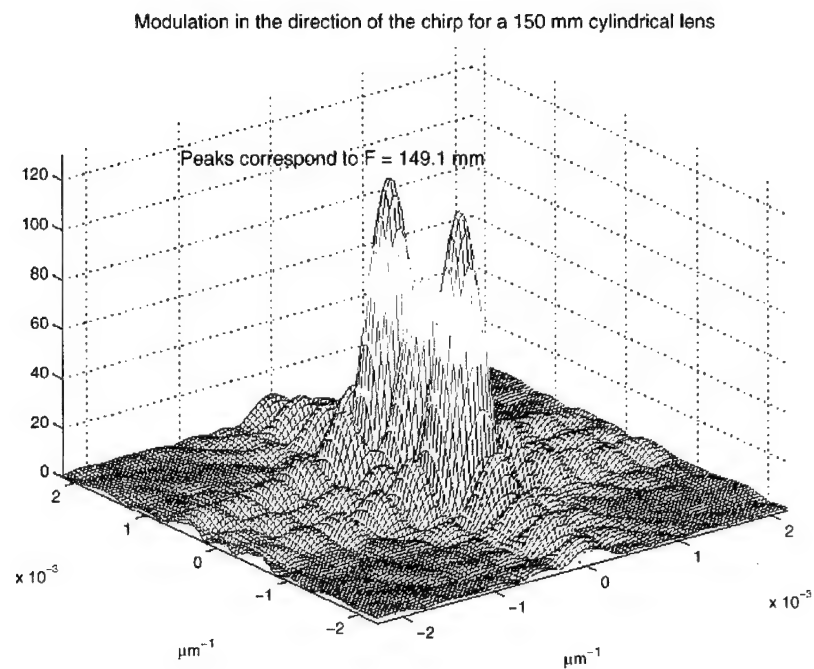
Figure 0.1   Measurement of a 150-mm cylindrical lens.

Modulation in the direction of the chirp for a 100 mm cylindrical lens

Peaks correspond to F = 97.87 mm

Modulation in the direction perpendicular to chirp for a 100 mm cylindrical lens

Peak corresponds to F = ∞
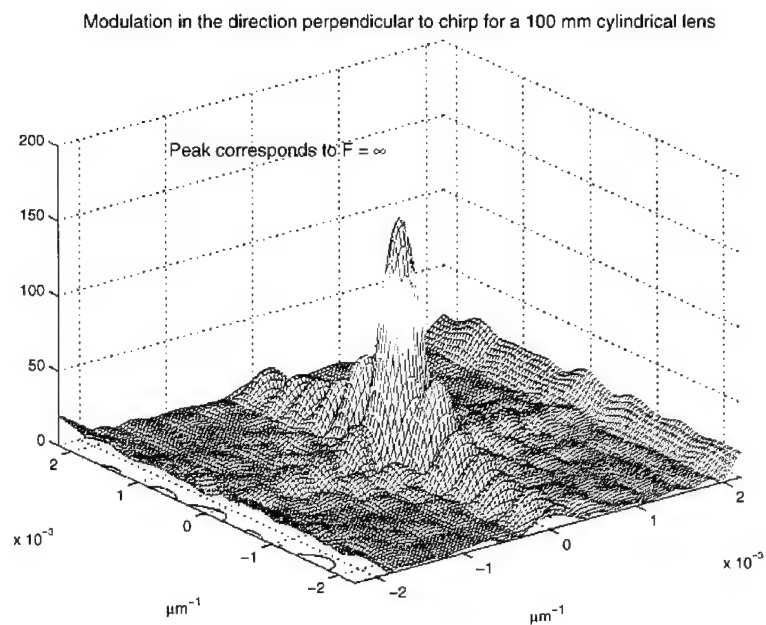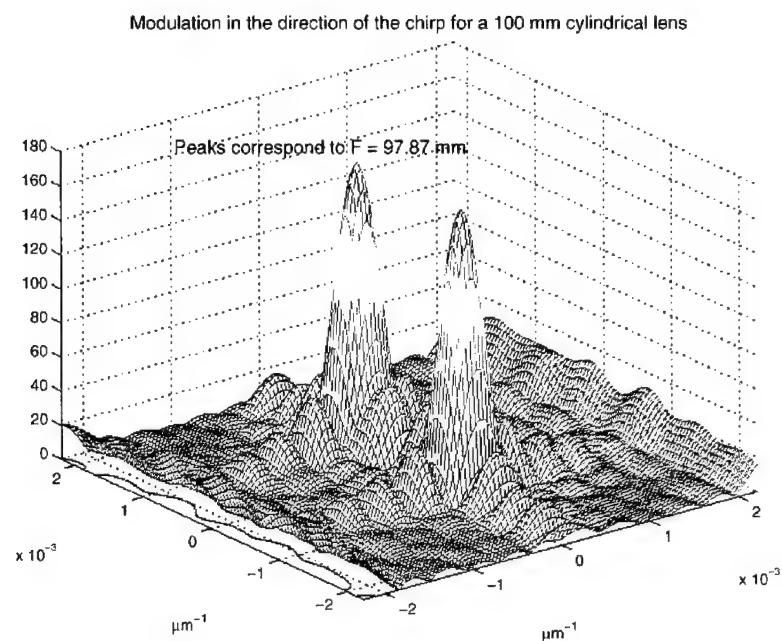
Figure 0.2   Measurement of a 100-mm cylindrical lens.

The output of the system also provided diverse phase shift data according to the theory developed in Chapter 3. As shown in Chapter 3, the data in different output pixels (that detect shear along the same direction) can be phase-shifted by choosing the right sequence of birefringent plates (see Eq. (3.19) in Chapter 3). In addition, the pixels in the characteristic pattern accumulate a π phase shift as soon as the wavefront of their respective fields passes through focus [2]. However, the π phase shift does not provide phase shift diversity. Only one of the two characteristic patterns suggested in Chapter 3 actually provided phase shift diversity. Unfortunately, as discussed in Chapter 4, the design of the custom plate holders prevented us from aligning the plates with accuracy better than $5°$, which translated into 20% variation of the output fan-out coefficients (see Tables 4.1 and 4.3). Each of the output phase pixels that provided phase shift diversity (pixels labeled 7, 10, 2, and 15 in Figure 3.5) was surrounded by four output phase pixels that did not provide phase shift diversity. Cross-talk and the finite dynamic range of the data from the CCD array prevented us from obtaining as good a signal in the phase-shifted pixels as in the non-phase-shifted ones. Figure 0.3 shows the interference signals from a phase-shifted pixel compared to the signal from one of the two pixels without phase shift. Although the signal in the phase-shifted pixel was strong enough to estimate the phase shift, it had a much lower SNR compared to the signal in the non-phase-shifted pixels (power SNR 0.63 compared to 6.29) and it was not considered in the sequence of experimental tests described in Chapter 4. In fact, Chapter 4 described experimental results obtained with a different pattern which provided better alignment but lacked phase diversity altogether.

The fact that the experiments clearly confirmed the theoretical calculations and that phase shifted output was obtained (albeit a poor SNR) leads us to conclude that the designed system is capable of outputting complete phase-shifted interferometric data. The signal in the phase-shifted pixels can be improved by simply using plate holders that allow a more accurate alignment of the system and by decreasing the cross-talk to even lower levels through better imaging of the input plane to the output.
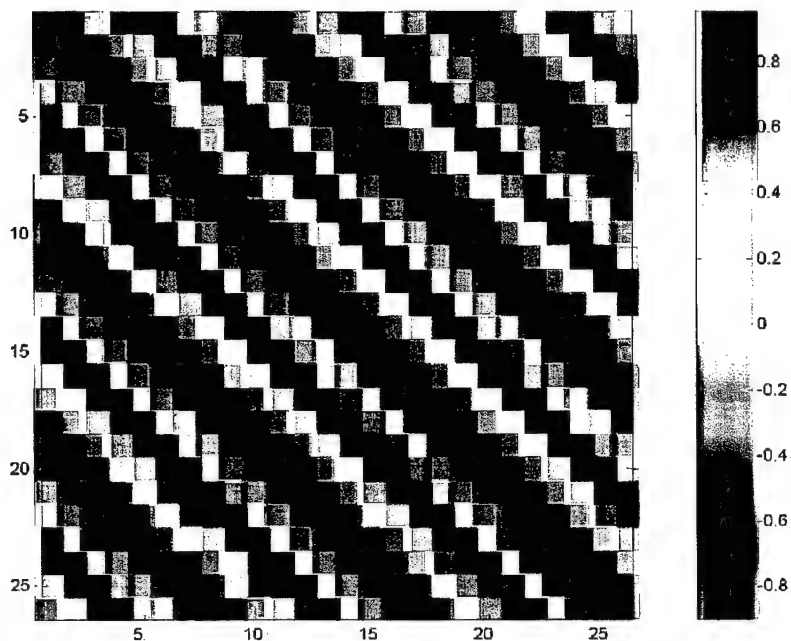
(a) Signal in a non-phase-shifted pixel.
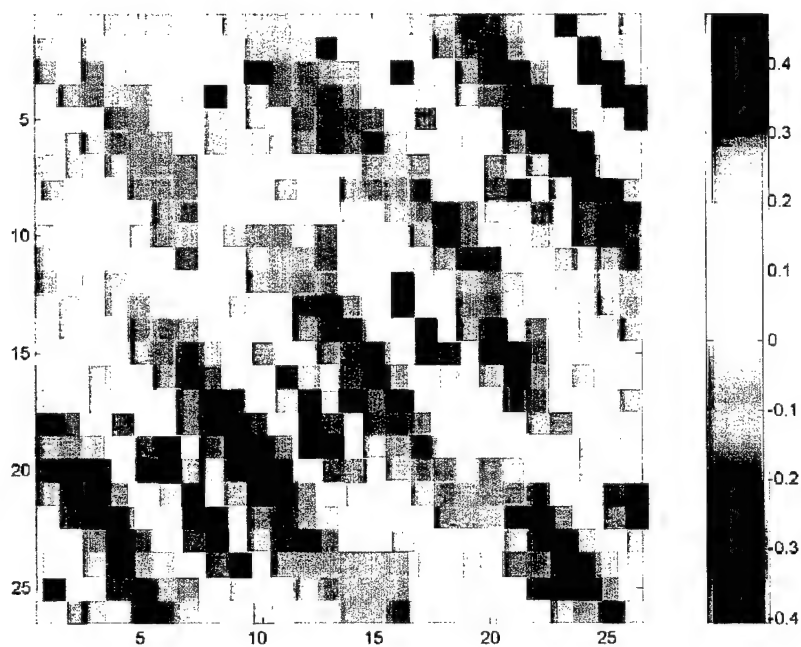
(b) Signal in a phase-shifted pixel.

Figure 0.3   Output signals from phase pixels with and without phase shift for pattern (a) in Figure 3.5.

We tested the system's sensitivity to vibrations in the environment, this being the main limitation of interferometric systems, which prevents their more widespread use. We caused vibrations of the optical table, which produced more than one λ change (RMS error) in the internal phase difference of a Mach-Zehnder interferometer. By comparison, our system showed a maximum of $\lambda/100$ RMS error with some of the pixels showing $\lambda/700$ RMS error. Our system is, therefore, at least 100 times less sensitive to vibrations than a classical system. This is due mainly to its common-path design. Also, we measured the internal phase error of the system without vibrations and obtained an RMS of $\lambda/3000$, comparable to that of high-performance phase shift interferometric systems. It is remarkable that this performance comes from an interferometric system laid out on a 1-m path and without any shielding from air currents.

Finally we measured the accuracy of the phase estimation by measuring minute phase changes produced by shifting laterally the convergent wavefront from the laser aplanat. We measured an accuracy of $\lambda/200$, comparable to the internal phase error of the system under environmental vibration noise. This was due to the fact that moving the input beam caused the speckle pattern on the CCD to change in a manner similar to the case when the system was subjected to environmental noise. Speckle and multiple interference noise were found to be the main noise sources in the system. This was shown by a decrease in noise of only 50% in the K parameters for the no-vibration case compared to the case with vibration noise. Further, immediate improvements of the system by making it even more compact, reducing the number of reflections inside the set of plates, and using a lower coherence laser source can minimize the speckle and multiple-interference noise. Thus, even better accuracy can be obtained.

Fundamental Parameters and Limitations

In this section we discuss some of the fundamental characteristics/limitations of the SFS. We compare the two implementations and show that imaging the input plane to the output improves the system dramatically.

> *Throughput*

The throughput is in turn given by

$$T_r = F_r \times \frac{2 \times NA_{sampling}}{\Omega} \qquad (1.11)$$

where F_r is the sampling fill ratio, $NA_{sampling}$ is the sampling numerical aperture of the system, i.e., half of the angle subtended by the detector element as viewed from the sampling plane, and Ω is the angular bandwidth of the sampled field. Fourier optics shows that Ω is given by $2\lambda/a$.

> *Fill ratio*

The fill ratio is given by $(a/h)^2$, with a being the sampling hole size and h being the sampling period. For the diffractive SFS, the fill ratio was given by the limit when the output field could not be described in the chosen manner, as a superposition of receiver patterns, with coefficients being the complex field at the sampling points. It was found to be approximately 1/100. For the imaging SFS (birefringent fan-out), the fill ratio is that of a regular shearing interferometer, such as a Mach-Zehnder. We have obtained good contrast fringes (see figures in Chapter 4) for an input wavefront with a radius of curvature of 180 mm, sampled approximately 4 mm off axis, i.e., having an equivalent local spatial frequency of $1/90\lambda$ (equivalent Nyquist sampling period of 45λ). The sampling hole size was approximately 15λ, giving a fill ratio of 1/9, 11 times better than that of the diffractive SFS.

> *Sampling numerical aperture*

The sampling numerical aperture of the diffractive SFS is given by l_d/d, where l_d is the dimension of the detector element and d is the distance to it. The size of the detector element was 4 μm and the distance to it was 1400 μm, thus giving $NA_{sampling}$ for the diffractive SFS to be 1/700. In the case of the imaging SFS, the sampling numerical aperture is actually the numerical aperture of the imaging system from the sampling plane to the output plane. The exact number would be obtained from ray tracing through the set of lenses. We used multielement commercial camera lenses and could not obtain designs for all of them, preventing us from performing the ray tracing. For the purpose of this calculation we chose

to underestimate this parameter at 0.1, which is also consistent with our accurate imaging of the 10-μm sampling holes through the system. Thus we obtained a 70-fold improvement by imaging the sampling plane to the output.

➢ *Sampling period*

The size of the sampling period is limited, in the case of the diffractive SFS, by the fill ratio and the sampling hole size. Sampling hole sizes less than 5 μm will cause the sampled field to depend on the incident field via a vector diffraction model, not considered in our approach. In order to avoid this, the sampling period needs to be bigger than approximately 50 μm and the hole size bigger than 5 μm. In the case of the imaging SFS, using the same sampling hole size (5 μm) but the 1/3 ratio of the size of the hole to the sampling period, gives a minimum sampling period of approximately 15 μm.

➢ *Space bandwidth product*

The SBP is given by

$$SBP = 2\frac{L_x L_y}{4h_x h_y} \tag{1.12}$$

For equal sampling hole sizes and the same input apertures, the imaging SFS has 9 times the SBP of the diffractive SFS, due to the higher fill ratio (the sampling period can be three times smaller).

➢ *SNR*

The most important parameter is the SNR of the data. Neglecting the speckle and multiple interference noise, it is given by

$$SNR = \frac{S}{N} = \frac{I_d \cdot QE \cdot G \cdot T}{\sqrt{N_R^2 + N_D^2 + N_S^2}} \tag{1.13}$$

where S is the total signal and N is the total noise in the measurements, I_d is the flux of photons incident on each detector element, QE is the quantum efficiency of the device, G is the pre-amplifier gain, T is the integration time, N_R is the read-out noise, N_D is the dark

charge noise, and N_S is the shot noise. If the read-out and the dark current noise terms are neglected (typical for light saturated applications) the SNR becomes

$$SNR = \frac{S}{N} = \frac{I_d \cdot QE \cdot G \cdot T}{\sqrt{I \cdot QE \cdot T}} = G\sqrt{I_d \cdot QE \cdot T} \qquad (1.14)$$

The SNR is therefore increasing as the square root of the photon flux reaching each detector element. The flux reaching each detector element is equal to $T_r \times I_{in}$, where T_r is the throughput and I_{in} is the input flux of photons reaching each sampling period. For the same input flux per sampling period the ratio of the SNRs is given by the ratio of the throughputs of the two sensor implementations. By substituting the fill ratio and the sampled field bandwidth in Eq (1.11) we obtain that

$$T_r = \frac{a^3}{h^2} \times \frac{NA_{sampling}}{\lambda} \qquad (1.15)$$

The ratio of the throughputs of the two sensors for the same sampling period is thus given by

$$\frac{T_r(\text{imaging SFS})}{T_r(\text{diffractive SFS})} = \left(\frac{a_{imaging}}{a_{diffractive}}\right)^3 \times \frac{NA_{sampling;imaging}}{NA_{sampling;diffractive}} \qquad (1.16)$$

The first ratio is 3 while the second is 70, from above. Therefore, the ratio of the throughputs is 1890. Using Eq. (1.14), this gives 44 times improvement in the SNR.

References

[1]   T. W. Stone and J. M. Battiato, "Optical array generation and interconnection using birefringent slabs," *Applied Optics*, vol. 33, pp. 182-91, 1994.

[2]   M. Born and E. Wolf, *Principles of Optics*, 6th ed. New York: Pergamon Press, 1980.

Wave-front sensing with a sampling field sensor

Remy Tumbar, Ronald A. Stack, and David J. Brady

We present a new type of optical wave-front sensor: the sampling field sensor (SFS). The SFS attempts to solve the problem of real-time optical phase detection. It has a high space-bandwidth product and can be made compact and vibration insensitive. We describe a particular implementation of this sensor and compare it, through numerical simulations, with a more mature technique based on the Shack–Hartmann wave-front sensor. We also present experimental results for SFS phase estimation. Finally, we discuss the advantages and drawbacks of this SFS implementation and suggest alternative implementations. © 2000 Optical Society of America

OCIS codes: 120.3180, 120.3930, 120.4640, 120.5050, 350.4600, 350.5030.

1. Introduction

Wave-front sensing is used to detect the amplitude and phase of the field for optical testing and hybrid optical imaging systems.[1] Optical testing applications include surface reconstruction techniques for eye surgery,[2] lens testing,[3] and diffractive optical element characterization.[4] Hybrid imaging systems use amplitude and phase measurements for postdetection or predetection processing. Adaptive optical systems[5] use predetection processing to improve image quality by removing wave-front distortions. Deconvolution[1] and tomographic reconstruction[6] systems use postdetection processing for digital object reconstruction.

Wave-front sensors include interferometric and noninterferometric systems.[7] Shearing interferometers, point-diffraction interferometers,[8] and the pseudo-phase-conjugate interferometer[9] are a few examples of interferometric wave-front sensors. Noninterferometric wave-front sensors include the Shack–Hartmann sensor (SHS)[10] and the curvature sensor[11] with the SHS being the most commonly used wave-front sensor. The SHS is constructed by one mounting an array of lenses in front of an array of detectors. The detectors determine the position of the focal spot intensity centroid of each lens. The set of centroid positions for a normally incident plane wave serves as a zero reference. The offsets between the detected centroids for an arbitrary input wave front and the zero reference positions provide measures of the average wave-front tilt coefficients over each lens subaperture. The input wave front is then reconstructed with data reduction techniques by use of these average tilt measurements.

In one technique we can find the wave-front phase at each sampling point (for example, the center of each lens) by considering that phase differences between adjacent sampling points are given by a linear combination of the local tilt coefficients. This technique is known as zonal reconstruction.[12–14] Zonal reconstruction has a low computational complexity because it uses sparse matrices to connect the values of the input wave front on the sampling grid to the measurements taken by the sensor. However, when the bandwidth of the input wave front is large, the higher-order aberration coefficients can no longer be neglected. In other words, the average tilt measurements are not a good representation of a high-bandwidth input field, and the zonal reconstruction method fails. In this case one could use a more complete representation of the subaperture wave front in terms of structure functions[15] or relate the average tilt measurements to a better representation of the input, which is the essence of modal reconstruction techniques.[14,16] The first approach needs more processing than the zonal reconstruction method, whereas the second one is faulted by modal cross coupling or aliasing.[17]

In this paper we present a new wave-front sensing device, the sampling field sensor (SFS). The SFS is a self-referencing interferometric wave front sensor similar, in this respect, to a shearing interferometer,

The authors are with the Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801. D. J. Brady's e-mail address is dbrady@uiuc.edu.

or to a point-diffraction interferometer. The SFS consists of a sampling stage and a fan-out stage. The input wave front is sampled with an array of small apertures. Each sample of the field is fanned out to multiple photodetectors in the sensor's output plane, where it overlaps with the field fanned out from adjacent samples. The interference from each pair of samples is detected, with different phase shifts, at a set of points in the output plane associated with the pair. The phase difference between the field coming from adjacent samples is estimated by use of techniques similar to phase-shift interferometry. In this paper we test numerically and experimentally a SFS design in which the sampling stage is a mask patterned with holes and the fan-out stage is simply Fresnel diffraction. We also show how a SFS can be built by placing an array of tapered sampling holes a small distance in front of a CCD array. This results in a compact, vibration-insensitive interferometric sensor. It combines the sensitivity of interferometric wave-front sensors with the robust, compact nature of the noninterferometric ones. In addition, we suggest alternative implementations of the fan-out stage. We use a numerical simulation to show that the SFS could sense input fields with bandwidths up to the Nyquist limit using only zonal reconstruction methods. This makes the SFS a good candidate for applications that require detection of high information content fields, i.e., fields with a high space-bandwidth product.[18]

In Section 2 we provide the SFS design equations and the input reconstruction (inversion) algorithm. In Section 3 we choose a particular set of design parameters and test this particular design numerically and experimentally. We discuss the results of the tests and propose alternative designs for the sampling and fan-out stages. In Section 4 we outline the characteristics of the SFS that could help spread use of optical phase detection systems outside the laboratory environment.

2. Sampling Field Sensor Design

A. General Principle

The principle behind the SFS can be viewed independently of the particular physical implementation of its constitutive elements. It relies on the fact that a band-limited function, such as an optical field, can be approximately represented over a finite aperture by a finite number of sample values. The SFS uses optical elements to separate sample values from the input field and to interfere them such that their phase and amplitude can be decoded. In the particular implementation we describe here we sample the input field using a hole grating and then detect the irradiance a specific distance away. This distance is chosen to ensure that the adjacent far-field patterns from each sampling hole overlap partially at the edges in the detection plane. Figure 1(a) illustrates the concept by showing only two sampling holes and their partially overlapping far-field diffraction patterns. We processed multiple intensity measurements in each overlap region using an algorithm
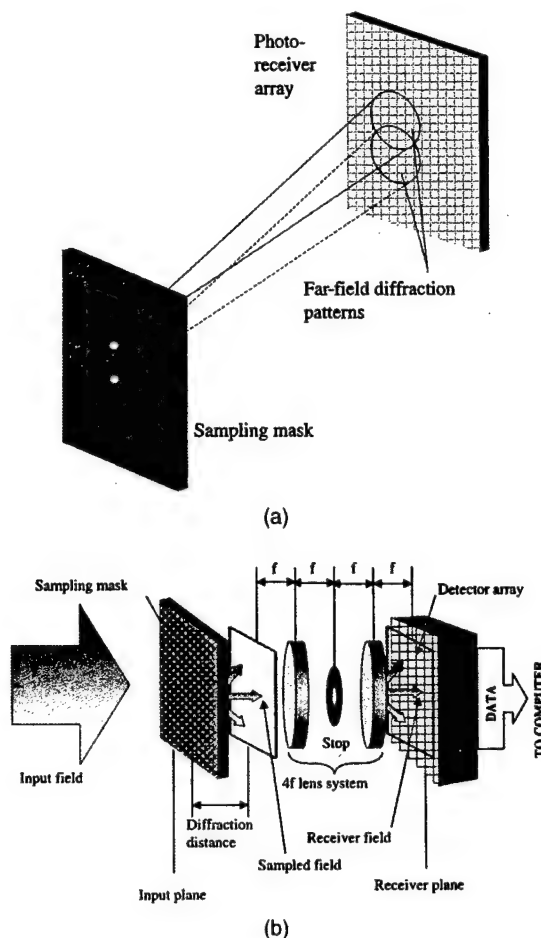
Fig. 1. (a) SFS concept. (b) Physical implementation of the SFS.

similar to phase-shifting interferometry to estimate the phase difference between the two adjacent sampling points. One measurement in each nonoverlap region is enough to estimate the irradiance of the corresponding sampling point. We can obtain the phase shifts required by the phase estimation algorithm by placing the detectors at specified positions in the output plane to take advantage of the quadratic phase shift given by the propagation in free space. In practice, one also includes spectral or spatial filtering components or other types of focusing or windowing components in the sensor.

B. Design Procedure

The goal of the SFS is to estimate the phase and amplitude of a band-limited field across an input aperture. We consider an input field $f_{in}(x, y)$ with bandwidth B. According to the Whittaker–Shannon theorem,[19]

$$f_{in}(x, y) = \sum_{m,n} f_{in}(mh, nh)\text{sinc}(x - mh)\text{sinc}(y - nh).$$

$$(1)$$

Equation (1) is valid under the Nyquist condition

$$h \leq \frac{1}{2B}.$$

$$(2)$$

If we could measure the field directly, we could use Eq. (1) to determine $f_{\text{in}}(x, y)$ by measuring sample values. In optical systems, however, one measures the time average of the intensity of the field rather than the field itself. In general, reconstructing $f_{\text{in}}(x, y)$ from samples of $\langle |f_{\text{in}}(x, y)|^2 \rangle$ is difficult or impossible. The idea of the SFS is to use sampling and linear transformation between the input aperture and the intensity at the detection plane to obtain an invertible relationship between the detected intensity and the input field. Sampling, by use of pinholes, for example, isolates the field values $f_{\text{in}}(mh, nh)$. Diffraction or other linear transformations between the pinholes and the detection plane maps the field samples onto new distributions. The field at the detection plane is then represented as

$$f_{\text{receiver}}(x', y') = \sum_{m,n} f_{\text{in}}(mh, nh)r(x' - mh', y' - nh'),$$

$$(3)$$

where $r(x, y)$ are called receiver pattern functions. We seek to design the receiver pattern $r(x, y)$ such that the field values $f_{\text{in}}(mh, nh)$, and thus $f_{\text{in}}(x, y)$, can be estimated from samples of $\langle |f_{\text{receiver}}(x, y)|^2 \rangle$. Practical difficulties that one confronts in implementing this design include the fact that Eq. (3) is not quite accurate because it is not possible to isolate $f_{\text{in}}(mh, nh)$ exactly with finite pinholes and the fact that reconstruction from $\langle |f_{\text{receiver}}(x, y)|^2 \rangle$ will be sensitive to noise. Discussion of these practicalities is deferred to Section 3 because in this section we are more concerned with the physical implementation of the SFS design concept.

Figure 1(b) shows a more detailed diagram of the SFS. In the input plane of the SFS there is a sampling mask consisting of small (compared to the sampling distance) clear regions or holes on an otherwise opaque substrate. Each hole is circular with diameter a and is placed on a rectangular grid with spacing h in both directions. This is the sampling stage of the device. The fan-out stage consists of a $4f$ system with its object plane placed a distance d behind the sampling mask. The Fourier plane of the $4f$ system has a low-pass filter with band limit B_L. The intensity in the output field of the $4f$ system is detected with an array of photodetectors, like a CCD camera. The output plane of the $4f$ system is the receiver plane shown in Fig. 1(b).

We now show how one can design the system described above to implement controlled fan-out from the sampling plane to the receiver plane and how to use intensities measured by the photodetectors in the receiver plane to estimate the amplitude and phase of the input field. The idea is to choose the parameters a, d, h, and B_L such that the field in the receiver plane of the $4f$ system can be considered to have the functional form of Eq. (3) with an appropriate receiver pattern. We consider a receiver pattern to be appropriate if Eq. (3) can be efficiently and accurately inverted to estimate the field samples. To achieve this

goal, we design the SFS to have receiver patterns such that

$$|r(x', y')| = \begin{cases} > 0, \text{ for } 0 \leq |x'|, |y'| < h' - \varepsilon/2 \\ = 0, \text{ for } |x'|, |y'| \geq h' - \varepsilon/2 \end{cases}. \quad (4)$$

Equation (4) implies that only two adjacent receiver pattern functions overlap at any point in the output plane. Also, there is a region in each receiver pattern of approximate size ε where there is no overlap. It is clear that the representation of the receiver field given by Eqs. (3) and (4) is an approximation. Because its output is band limited it cannot be space limited as suggested by Eq. (4). However, if the intensity in the receiver pattern function outside the specified support region is small enough, we can consider it to be zero. Therefore we specify the design parameters a, d, h, and B_L such that the error in Eq. (3) is negligible when $r(x, y)$ is of finite support.

The design equations are found by one considering the diffraction of the field scattered by the sampling mask using Kirchhoff boundary conditions (KBC) with a propagation kernel derived from a Fourier optics approach. Assuming the four design parameters have been found, we can describe the output of the system, to a good approximation, by Eq. (3). The next step is to then specify the positions of the photodetectors in the receiver plane. We assume that, in general, the placement of photodetectors or receiver pixels in the receiver plane can be arbitrary and is not limited to a regular square grid. Therefore these positions are additional design parameters. The final step is to specify an inversion algorithm to estimate the amplitude and phase of the input field at the sampling points based on the intensity measurements at the receiver pixels. We show in this paper that the measured intensity values depend on the amplitude and phase of the input field at each sampling point by way of a transformation that is parameterized by a set of constants. The constants depend on the construction of the SFS, especially the position of the receiver pixels (x_p, y_p). We choose these positions such that the resulting set of parameters will give a robust inversion of the data to reconstruct the input.

To specify the design parameters a, d, h, B_L, and (x_p, y_p) we must consider the physical implementation of the SFS in more detail. Assuming KBC, the field right after the sampling mask $f_{\text{sampled}}(x, y)$ is given by

$$f_{\text{sampled}}(x, y) = f_{\text{in}}(x, y) \sum_{j,k} s_0(x - jh, y - kh)$$

$$= \sum_{m,n} f_{\text{in}}(mh, nh)\varphi(x - mh, y - nh), \quad (5)$$

where the function $s_0(x, y)$ describes the transmission through one period of the mask and is given by

$$s_0(x) = \begin{cases} 1 \text{ for } |x|, |y| \leq \dfrac{a}{2} \\ 0 \text{ for } \dfrac{a}{2} < |x|, |y| < \dfrac{h}{2} \end{cases}. \quad (6)$$

The function $\varphi(x, y)$ is

$$\varphi(x, y) = \psi(x, y) \sum_{j,k} s_0(x - jh, y - kh), \qquad (7)$$

where $\psi(x, y)$ is $\mathrm{sinc}(x, y)$. The receiver field is the sampled field convolved with the free-space propagation Fresnel kernel, followed by the convolution with the point-spread function of the band-limited $4f$ system. Thus the receiver field is represented by

$$f_{\mathrm{receiver}}(x, y) = [f_{\mathrm{sampled}}(x, y) \otimes h_d(x, y)] \otimes h_{4-f}(x, y), \qquad (8)$$

where $\otimes$ is the two-dimensional convolution operator, $h_d(x, y)$ is the Fresnel kernel representing propagation over distance d, and $h_{4f}(x, y)$ is the point-spread function of the $4f$ system. Note that the consideration of the second convolution operation in Eq. (8) neglects vignetting.[19] We can also write Eq. (8) using a property of the Fourier transform as

$$\begin{aligned} f_{\mathrm{receiver}}(x, y) &= \mathrm{FT}^{-1}[F_{\mathrm{sampled}}(u, v) H_d(u, v) H_{4f}(u, v)] \\ &= \mathrm{FT}^{-1}[F_{\mathrm{sampled}}(u, v) H_{\mathrm{SFS}}(u, v)], \qquad (9) \end{aligned}$$

where FT^{-1} is the inverse Fourier-transform operator, H_d is the Fourier transform of h_d, H_{4f} is the Fourier transform of h_{4f}, and H_{SFS} is the transfer function of the SFS, i.e., the Fourier transform of the point-spread function of the SFS. If we assume that the $4f$ system has no aberrations and that it implements a low-pass filter in its Fourier plane, H_{4f} will be a low-pass filter also. Therefore H_{SFS} is the Fourier transform of the Fresnel kernel windowed by the low-pass filter in the Fourier plane of the $4f$ system. Using Eqs. (5)–(8) we find that the field in the receiver plane is

$$\begin{aligned} f_{\mathrm{receiver}}(x, y) &= f_{\mathrm{sampled}}(x, y) \otimes h_{\mathrm{SFS}} \\ &= \sum_{m,n} f_{\mathrm{in}}(mh, nh) r(x - mh', y - nh'), \qquad (10) \end{aligned}$$

where $h_{\mathrm{SFS}} = \mathrm{FT}^{-1}(H_{\mathrm{SFS}})$, and the receiver pattern function is given by

$$\begin{aligned} r(x - mh', y - nh') &= \varphi(x - mh', y - nh') \\ &\otimes h_{\mathrm{SFS}}(x, y) \qquad (11) \end{aligned}$$

over all integers m and n. The parameter h' is the distance between the centers of adjacent receiver patterns in the receiver plane. We can assume, without loss of generality, $1\times$ magnification for the $4f$ system. Thus h' is equal to h, and the coordinates in the receiver plane are the same as in the sampling plane. Because the sampling mask is periodic with period h, it follows that $r(x\text{-}mh, y\text{-}nh) = r(x, y)$ over all integers m and n. Substituting Eq. (7) into Eq. (11) yields the receiver pattern function

$$r(x, y) = \left[\psi(x, y) \sum_{j,k} s_0(x - jh, y - kh) \right] \otimes h_{\mathrm{SFS}}(x, y). \qquad (12)$$

The limited support requirement in Eq. (4) cannot be satisfied, in general, for the case in which KBC apply and h_{SFS} is band limited because both the factor in the square brackets in Eq. (12) as well as h_{SFS} have unlimited support. However, if the size of the sampling hole is small compared to the sampling distance, the contributions from the zero crossings of $\psi(x, y)$ can be neglected. This allows the factor in the square brackets in Eq. (12) to satisfy the requirement of limited support. The size of the sampling hole can be found by one trying different values until the SFS wave-front reconstruction error becomes appropriately small, i.e., when the expansion in Eq. (3) is still valid with the limited support requirement of Eq. (4) satisfied. Assuming that we can neglect the terms in the sum that multiply the regions around the zero crossings of $\psi(x, y)$ and allow $\psi(x, y) \cong 1$ over the sampling hole region in the main lobe of $\psi(x, y)$, we use Eq. (12) and the definition of h_{SFS} to obtain

$$\begin{aligned} r(x, y) &= [\psi(x, y) s_0(x, y)] \otimes h_d(x, y) \otimes h_{4f}(x, y) \\ &\cong s_0(x, y) \otimes h_{4f}(x, y) \otimes h_d(x, y). \qquad (13) \end{aligned}$$

By grouping the two factors of the first convolution in $s_0'(x, y)$ and using the Fresnel diffraction kernel[19] for h_d, we obtain

$$\begin{aligned} r(x, y) &\cong \exp\left[j \frac{(x^2 + y^2)}{\lambda d} \right] \mathrm{FT}\bigg\{ s_0'(\xi, \eta) \exp \\ &\times \left[j \frac{(\xi^2 + \eta^2)}{\lambda d} \right] \bigg\} \bigg|_{u = \frac{x}{\lambda d}, v = \frac{y}{\lambda d}}, \qquad (14) \end{aligned}$$

where u and v are the spatial frequencies where the Fourier transform is evaluated. Thus, for a sampling hole of size a much smaller than h and d, the receiver pattern function is approximately given by the Fourier transform of the band-limited sampling hole $s_0'(x, y)$ with an additional quadratic phase factor. We can pick the windowing function in the Fourier plane of the $4f$ system such that it reduces the ripples of h_{SFS} outside of the specified support region. If we pick an appropriate windowing function (Hanning or Hamming) with a band limit

$$B_L = \frac{1}{a}, \qquad (15)$$

then $r(x, y)$ will be given only by the main lobe of a sinc function.

In Section 3.A we give more details about the numerical simulations of the SFS wave-front reconstruction. Among other tests, we considered a sinc-shaped incident field diffracted by a mask in which we sampled the input function at the central lobe and adjacent zero crossings. This gives an output field $r(x, y)$ that consists of a set of diffraction patches arranged on a grid with spacing h. We observed that a ratio h/a of 10 makes the ratio of the intensities in the patches coming from the zero crossings and the central patch to be of the order of 10^{-4}. This ratio increases with our increasing the size of the hole

compared to the sampling distance. Also, the case with $h = 110$, $a = 10$, $\lambda = 0.633$, $d = 1400$ μm and a Hanning window with $B_L = 0.1$ μm^{-1} gave a 20% mean reconstruction error at maximum input band-width. Considering this to be satisfactory, we obtained a design equation giving a to be roughly

$$a \approx \frac{h}{10}. \tag{16}$$

Enforcing the limited support of the receiver pattern functions in the context of using KBC to model the diffraction from the sampling mask severely limits the ratio a/h. The main drawback is to severely limit the amount of light transmitted by the system. We also note that the model presented here is not valid for h less than approximately 70 μm because this would require a less than 7 μm or approximately 10 wavelengths, which in turn makes the KBC invalid. We can derive the equation that gives the propagation distance d from relation (14) by noting that $s_0'(x, y)$ is band limited to B_L and that the support of $r(x, y)$ is required to be inside the square $|x| \le qh$ and $|y| \le qh$ with $q < 1$. On the other hand, d needs to be large enough to ensure the overlap of adjacent patterns, i.e., $|r(x, y)|$ has to be nonzero for $|x| \ge h/2$ and $|y| \ge h/2$. Thus we obtain

$$\frac{h}{2\lambda B_L} \le d \le \frac{qh}{\lambda B_L}. \tag{17}$$

To summarize, Eqs. (2) and (15), approximation (16), and inequality (17) are the design equations that give the parameters h, B_L, a, and d, respectively.

Next we consider the equations giving the positions of the receiver pixels, but first we need to model their intensity measurements. We specify the detection area of each receiver pixel by the function $\sigma_{i,j}(x_p, y_p)$, where the individual receiver pixel coordinates are denoted by x_p and y_p with respect to the center of the corresponding receiver pattern $r(x-ih, y-jh)$, which is centered at (ih, jh). The function $\sigma_{i,j}(x_p, y_p)$ is unity over the area of the detector centered at (x_p, y_p) in receiver cell (i, j) and zero elsewhere. The receiver cell (i, j) is the region of support for $r(x-ih, y-jh)$. Thus the signal measured by the photodetector at the receiver pixel $\sigma_{i,j}(x_p, y_p)$ is proportional to

$$\Gamma_{i,j}(x_p, y_p) = \iint\limits_{\sigma_{i,j}(x_p, y_p)} \left| \sum_{m,n} f_{\text{in}}(mh, nh) r(x'' - mh', \right.$$
$$\left. y'' - nh') \right|^2 dx'' dy'', \tag{18}$$

where at most two terms of the sum interfere at each point according to the finite support requirement of Eq. (4). We consider three classes of receiver pixels: amplitude pixels, phase 1 pixels, and phase 2 pixels. The amplitude pixels are placed in the nonoverlap regions at the center of each receiver pattern. Therefore the size of the amplitude pixels needs to be

less than ε or $2h(1-q)$ in each direction. Both types of phase pixels are placed in the overlap regions. All three classes of receiver pixels are placed in the receiver plane in a periodic structure. Each receiver cell will contain nine receiver pixels: one amplitude pixel, four phase 1 pixels, and four phase 2 pixels. Figure 2 shows three adjacent, overlapping receiver cells. The signal in the amplitude pixel of receiver cell (m, n) is given by

$$\Gamma_{m,n}(0, 0) = |f_{\text{in}}(mh, nh)|^2 \iint\limits_{\sigma_{m,n}(0, 0)} |r(x'' - mh',$$
$$y'' - nh')|^2 dx'' dy''. \tag{19}$$

Equation (19) can be written in the form

$$\Gamma_{m,n}(0, 0) = |f_{\text{in}}(mh, nh)|^2 K_a^2, \tag{20}$$

where K_a is a real number that is independent of the input, depending only on the construction of the SFS. The signal measured by the phase pixels differs from Eq. (19) in that the overlap of two adjacent (horizontally or vertically) receiver patterns needs to be considered. Thus the signal for a phase-type receiver pixel in the receiver cell (m, n) is found by one keeping two horizontally (vertically) adjacent terms from Eq. (18). This results in

$$\Gamma_{m,n}(x_{Pi}, y_{Pi}) = |f_{\text{in}}(mh, nh)|^2 \iint\limits_{\sigma_{m,n}(x_{Pi}, y_{Pi})} |r(x'' - mh',$$
$$y'' - nh')|^2 dx'' dy'' + |f_{\text{in}}[(m + 1)h, nh]|$$
$$\times \iint\limits_{\sigma_{m,n}(x_{Pi}, y_{Pi})} |r[x'' - (m + 1)h',$$
$$y'' - nh']|^2 dx'' dy''$$
$$+ 2 \operatorname{Re}\{f_{\text{in}}(mh, nh)^* f_{\text{in}}[(m + 1)h, nh]\}$$
$$\times \iint\limits_{\sigma_{m,n}(x_{Pi}, y_{Pi})} \operatorname{Re}\{r(x'' - mh',$$
$$y'' - nh')^* r[x'' - (m + 1)h',$$
$$y'' - nh']\} dx'' dy'', \tag{21}$$

where (x_{Pi}, y_{Pi}) is the position of each phase pixel of type $i = 1, 2$ in the x direction (note that the symbol * represents complex conjugate). Equation (21) can be rewritten as

$$\Gamma_{m,n}(x_{Pi}, y_{Pi}) = |f_{\text{in}}(mh, nh)|^2 K_{1i} + |f_{\text{in}}[(m + 1)$$
$$\times h, nh]|^2 K_{2i} + K_{Pi}|f_{\text{in}}(mh, nh)|$$
$$\times |f_{\text{in}}[(m + 1)h, nh]|\cos$$
$$\times \{\phi(mh, nh) - \phi[(m + 1)h, nh] - \alpha_i\}, \tag{22}$$
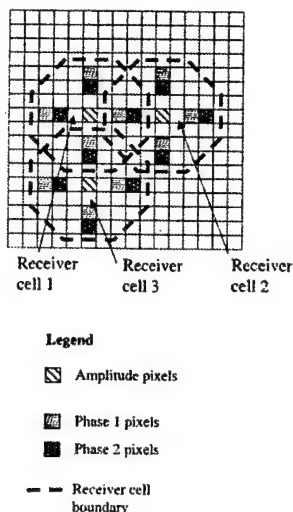
Fig. 2. Layout of overlapping adjacent receiver patterns. A receiver cell is the area covered by the corresponding receiver pattern.

where $\phi(x, y)$ is the phase of the complex wave front $f_{\text{in}}(x, y)$. Equation (22) also defines four other constants, K_{1i}, K_{2i}, K_{Pi}, and α_i for each phase pixel of type $i = 1, 2$. Thus there are a total of nine constants that define the output of the SFS system independently of the input, as long as the input is band limited and Eq. (3) with the restriction of Eq. (4) remains valid. These constants are K_a, K_{P1}, K_{P2}, K_{11}, K_{12}, K_{21}, K_{22}, α_1, and α_2. They are enumerated as follows:

$$K_a = \left[\iint_{\sigma_{0,0}(0, 0)} |r(x'', y'')|^2 dx'' dy'' \right]^{1/2},$$

$$K_{Pi} = \left| \iint_{\sigma_{0,0}(x_{Pi}, y_{Pi})} r(x'', y'')^* r(x'' - h, y'') dx'' dy'' \right| \text{ for } i$$
$$= 1, 2,$$

$$\alpha_i = -\angle \left[\iint_{\sigma_{0,0}(x_{Pi}, y_{Pi})} r(x'', y'')^* r(x'' - h, y'') dx'' dy'' \right] \text{ for } i$$
$$= 1, 2,$$

$$K_{1i} = \iint_{\sigma_{0,0}(x_{Pi}, y_{Pi})} |r(x'', y'')|^2 dx'' dy'' \text{ with } i = 1, 2,$$

$$K_{2i} = \iint_{\sigma_{0,0}(x_{Pi}, y_{Pi})} |r(x'' - h', y'')|^2 dx'' dy'' \text{ for } i = 1, 2.$$

(23)

Extending Eqs. (21)–(23) to the y direction is just a matter of considering the adjacency of the receiver

patterns in that direction and calculating the new constants. In certain cases, circularly symmetric or separable receiver patterns, the pixel assignment can be done so that the K parameters are the same in both Cartesian directions. By changing the values of m and n, one can relate the values of the field at all the points (mh, nh) to the signals detected in the receiver pixels by Eqs. (20), (22), and (23). Equations (20) and (22) constitute what we call the SFS representation of the input field, whereas Eqs. (23) give the physical meaning of its parameters. The inversion algorithm is based on the SFS representation, parameterized by the K parameters and the assumption that $\alpha_{1,2}$ are different. We directly estimate the amplitude at each sampling point from the intensity measurement at the amplitude pixel for each receiver cell using Eq. (20) with

$$|f_{\text{in}}^{\text{estimated}}(ih, jh)| = \frac{\sqrt{\Gamma_{i,j}(0, 0)}}{K_a}. \tag{24}$$

The phase difference between adjacent sampling points is found with a procedure similar to phase-shift interferometry.[20] The two intensities measured by both phase pixels of a receiver cell may be considered as two frames with different phase shifts in a phase-shift phase estimation algorithm. The difference is that the average intensity and the fringe modulation factor are changing from one phase pixel to another, making classical phase-shift techniques unusable. The alternative that we propose is to solve the system of equations formed by Eq. (22) with $i = 1, 2$ for the quantities

$$|f_{\text{in}}[(m + 1)h, nh]||f_{\text{in}}(mh, nh)|K_{p1}K_{p2} \sin\{\varphi(mh, nh)$$
$$- \varphi[(m + 1)h, nh]\} \sin(\alpha_1 - \alpha_2) = D_1 K_{p2} \cos(\alpha_2)$$
$$- D_2 K_{p1} \cos(\alpha_1)$$

$$|f_{\text{in}}[(m + 1)h, nh]||f_{\text{in}}(mh, nh)|K_{p1}K_{p2} \cos\{\varphi(mh, nh)$$
$$- \varphi[(m + 1)h, nh]\} \sin(\alpha_1 - \alpha_2) = -D_1 K_{p2} \sin(\alpha_2)$$
$$+ D_2 K_{p1} \sin(\alpha_1), \quad (25)$$

where

$$D_1 = \Gamma_{m,n}(x_{P1}, y_{P1}) - K_{11}|f_{\text{in}}(mh, nh)|^2 - K_{21}|f_{\text{in}}[(m + 1)h, nh]|^2,$$

$$D_2 = \Gamma_{m,n}(x_{P2}, y_{P2}) - K_{12}|f_{\text{in}}(mh, nh)|^2 - K_{22}|f_{\text{in}}[(m + 1)h, nh]|^2. \tag{26}$$

The values of D_1 and D_2 can be estimated by use of the values of the field amplitudes calculated with Eq. (20). Finally, we find the phase difference by taking the inverse tangent of the ratio of the two quantities in Eqs. (25). The result is extended over the $(-\pi, \pi)$ interval by use of the signs of the numerator and denominator. To summarize, the amplitude and phase of the input field are given by the following inversion algorithm:

$$|f_{\text{in}}^{\text{estimated}}(mh, nh)| = \frac{\sqrt{\Gamma_{m,n}(0, 0)}}{K_a},$$

$$D_1 = \Gamma_{m,n}(x_{P1}, y_{P1}) - K_{11}|f_{\text{in}}^{\text{estimated}}(mh, nh)|^2 - K_{21}$$
$$\times |f_{\text{in}}^{\text{estimated}}[(m + 1)h, nh]|^2,$$

$$D_2 = \Gamma_{m,n}(x_{P2}, y_{P2}) - K_{12}|f_{\text{in}}^{\text{estimated}}(mh, nh)|^2 - K_{22}$$
$$\times |f_{\text{in}}^{\text{estimated}}[(m + 1)h, nh]|^2,$$

$$\varphi^{\text{estimated}}(mh, nh) - \varphi^{\text{estimated}}[(m + 1)h, nh] = \tan^{-1}\left[\frac{D_1 K_{p2} \cos(\alpha_2) - D_2 K_{p1} \cos(\alpha_1)}{-D_1 K_{p2} \sin(\alpha_2) + D_2 K_{p1} \sin(\alpha_1)}\right]. \tag{27}$$

The phase map of the input wave front can be generated from these phase differences by use of well-known techniques from shearing interferometry. The last step of the derivation in this subsection is to give the rule for choosing the positions of the phase pixels (x_{P1}, y_{P1}) and (x_{P2}, y_{P2}). This has to be done such that it gives the best phase estimation error because the amplitude estimation depends only on K_a, therefore it is independent on the positions of the phase pixels. Dividing by $K_{P1}K_{P2}$ in the right-hand side of the last equation in Eqs. (27) shows that the phase estimation error depends on the estimation error of D_1/K_{P1} and D_2/K_{P2}. On the other hand, these depend on the amplitude estimation error through coefficients K_{11}/K_{P1}, K_{21}/K_{P1}, K_{12}/K_{P2}, and K_{22}/K_{P2}. Minimizing these coefficients improves the phase estimation error. To find the position of the phase pixels that will do this minimization, we use Eqs. (23) and approximate the receiver functions as triangle functions defined by

$$r(x, y) = \begin{cases} \left(1 - \dfrac{|x|}{qh}\right)\left(1 - \dfrac{|y|}{qh}\right) \text{ for } |x|, |y| \leq qh \\ 0 \text{ elsewhere} \end{cases}. \tag{28}$$

We also consider the regions of integration to be small enough such that we can approximate the integrals in Eqs. (23) with the values of the integrand at the receiver pixel position (x_{Pi}, y_{Pi}) or $(0, 0)$. Thus K_{11}/K_{P1} and K_{21}/K_{P1} are given by

$$\frac{K_{11}}{K_{P1}} = \frac{qh - x_{P1}}{(q - 1)h + x_{P1}},$$

$$\frac{K_{21}}{K_{P1}} = \frac{(q - 1)h + x_{P1}}{qh - x_{P1}}. \tag{29}$$

Similarly, K_{12}/K_{P2} and K_{22}/K_{P2} are given by

$$\frac{K_{12}}{K_{P2}} = \frac{qh - x_{P2}}{(q - 1)h + x_{P2}},$$

$$\frac{K_{22}}{K_{P2}} = \frac{(q - 1)h + x_{P2}}{qh - x_{P2}}. \tag{30}$$

From Eqs. (29) and (30) we find that, to obtain equal sensitivity to measurement errors in $|f_{\text{in}}^{\text{estimated}}(mh, nh)|$ and $|f_{\text{in}}^{\text{estimated}}[(m + 1)h, nh]|$ we need to have $x_{P1} = x_{P2} = h/2$ independently of y because of re-

ceiver pattern separability. A similar set of equations is found for y adjacency and, likewise, the result is that $y_{P1} = y_{P2} = h/2$ independently of x.

The last design constraint that needs to be considered is the phase difference $|\alpha_1 - \alpha_2|$. In fact, Eqs. (25) state that the sine and cosine of the phase difference depend on the numerator and denominator in the last equation of Eqs. (27) through $1/\sin(\alpha_1 - \alpha_2)$. If this is big then the errors in estimating the quantities in the numerator and the denominator will be amplified. Therefore we have the design equation

$$|\alpha_1 - \alpha_2| = (2p + 1)\frac{\pi}{2}, \tag{31}$$

where p is any positive integer or zero. To find the pixel positions that would satisfy the requirement placed on the difference between both phase angles, we use relation (14) and Eqs. (23). First we note that we can use the built-in output phase shift of the Fresnel transformation by appropriately positioning the receiver phase pixels. We consider two phase pixels placed between two adjacent amplitude pixels somewhere on the $y = 0$ line symmetrically with respect to the middle point $x = h/2$ that was found to be optimal for amplitude error sensitivity. The difference in phase angles is given by

$$\alpha_1 - \alpha_2 = \angle\left[\iint_{\sigma_{0,0}(x_{P1}, y_{P1})} r(x'', y'')^* r(x'' - h, y'') dx'' dy''\right]$$
$$- \angle\left[\iint_{\sigma_{0,0}(x_{P2}, y_{P2})} r(x'', y'')^* r(x'' - h, y'') dx'' dy''\right]. \tag{32}$$

Again, we assume the size of the phase pixels to be small enough such that to consider the integration over their areas as a multiplication by a delta function. By using relation (14) for the receiver pattern functions and neglecting the phase of the Fourier transform of $s_0'(x, y)$, we simplify Eq. (32) to obtain

$$|\alpha_1 - \alpha_2| = \left|\angle\left\{\exp\left[j\frac{h(2x_{P1} - h)}{\lambda d}\right]\right\} - \angle\left\{\exp\left[j\frac{h(2x_{P2} - h)}{\lambda d}\right]\right\}\right|. \tag{33}$$

Thus by substituting the requirement from Eq. (31) ($p = 0$) into Eq. (33) we obtain the design equation

$$|x_{P1} - x_{P2}| = \frac{\lambda d \pi}{4h}. \qquad (34)$$

Because Eq. (34) is found by drastically approximating Eqs. (23), it should thus be treated as a rule of thumb and used to obtain an order-of-magnitude value for the phase pixel separation distance. In fact, the actual phase pixel positions are found by trial and error iteration starting with values around the midpoint, $x = h/2$, and given by Eq. (34). Finally, because the intensity measured by the phase pixels is the integration of a local fringe pattern, the modulation of the detected signal is proportional to the ratio between the inverse of the local spatial frequency and the size of the receiver pixel. Therefore the modulation would increase with decreasing pixel size $\Delta x \times \Delta y$. However, this would decrease the level of the detected signal, and the biggest pixel size that still gives good reconstruction error has to be chosen. In Subsections 3.A and 3.B we test this SFS design both numerically and experimentally.

3. Testing the Sampling Field Sensor

A. Numerical Simulations

The following numerical simulations depend on particular design parameters chosen to closely match our experimental system. The particular design parameters are enumerated here for completeness as follows:

$$h = 110 \ \mu\text{m};$$

$$a = 10 \ \mu\text{m};$$

$$d = 1400 \ \mu\text{m};$$

$$B_L = 0.1 \ \mu\text{m}^{-1} \text{ with a Hanning window};$$

$$x_{P1x} = y_{P1y} = 55.0 \ \mu\text{m}, \ y_{P1x} = x_{P1y} = 0.0 \ \mu\text{m};$$

$$x_{P2x} = y_{P2y} = 55.0 \ \mu\text{m}, \ y_{P2x} = x_{P2y} = 7.0 \ \mu\text{m};$$

$$\Delta x = \Delta y = 3 \ \mu\text{m}. \qquad (35)$$

The meaning of the parameters in the last three equations in Eqs. (35) is shown in Fig. 3. Only one period of the periodic structure of receiver pixels is shown. The period of the structure is given to be $h = 110 \ \mu\text{m}$, and the focal lengths of the lenses in the $4f$ system are not given because we assumed a perfect unit magnification imaging system.

It is of interest to know how well the SFS described in Section 2 detects the amplitudes and the phase differences between adjacent samples of the input field. As a measurement of detection error we use the difference between the simulated input field (amplitude or phase), evaluated at the center point of each sampling period, and the estimate obtained us-
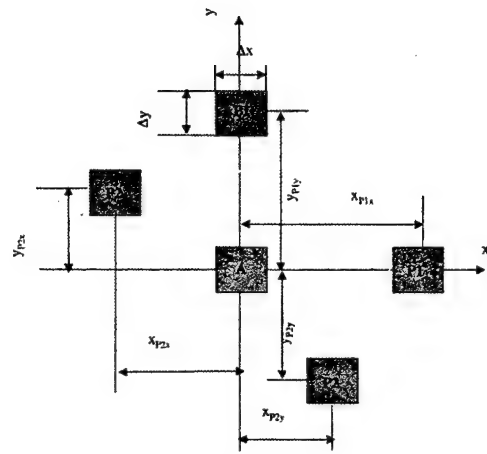
Fig. 3. Exact layout of the receiver mask, one receiver cell only. A marks the amplitude pixel and P1 and P2 are the two types of phase pixels.

ing our method and SFS device. That is, the detection error is defined as

$$\Delta f_{m,n} = \| f_{\text{in}}^{\text{estimated}}(mh, nh)| - |f_{\text{in}}(mh, nh)\|,$$

$$\Delta \varphi_{m,n;m,n+1} = |\{\varphi(mh, nh) - \varphi[mh, (n + 1)h]\}$$
$$- \{\varphi^{\text{estimated}}(mh, nh) - \varphi^{\text{estimated}}[mh, (n$$
$$+ 1)h]\}|, \qquad (36)$$

where Δf is the amplitude error and $\Delta \varphi$ is the phase error. Only x adjacency is shown in Eqs. (36) for brevity. We simulate the field transformation implemented by the SFS and described in Section 2 by considering it to be a windowed Fresnel transformation. As noted above, this neglects vignetting and the aberrations of the $4f$ system. The simulated input field was generated by approximate prolate interpolation[21] from a set of samples that oversampled the band-limited input field. In this way it is possible for one to go to a higher order of accuracy in the Fresnel diffraction calculations by changing only the interpolation distance and not the input samples. The samples had the real and imaginary parts uniformly distributed random variables in the interval $[-0.5, 0.5]$. The lateral dimensions of the simulated fields were 2.2 mm $\times$ 2.2 mm sampled at 1 μm in both Cartesian directions. We used fast Fourier transforms to calculate the Fresnel diffraction integral. This was performed in a serial fashion by our considering that the output of one receiver cell is due to the field coming from a corresponding 3 $\times$ 3 region in the input. The error would go to zero if the corresponding input region were extended to include all the input points. This would be impractical for the size of the input in our case. To check if the 3 $\times$ 3 window is big enough, we increased its size to 5 $\times$ 5 and ran a simulation for the field with the highest input bandwidth. We found that the change in the error was insignificant so we kept a 3 $\times$ 3 window in all subsequent tests. The SFS had 20 $\times$ 20 periods, hence there were 400 detection points. To calculate

Table 1. Error in Calibrating the SFS Parameters with a Least-Squares Fit[a]

Field Bandwidth (μm^{-1})	Phase Pixel 1		Phase Pixel 2	
	Percentile Error (%)	Residual Error[b] (a.u.)	Percentile Error (%)	Residual Error[b] (a.u.)
1/2200	0.26 ± 0.49	7.52×10^{-6}	0.19 ± 0.41	9.96×10^{-6}
1/1980	0.26 ± 0.46	8.26×10^{-6}	0.16 ± 0.6	8.40×10^{-6}
1/1760	0.19 ± 0.4	4.39×10^{-6}	0.14 ± 0.4	1.20×10^{-5}
1/1540	0.2 ± 0.35	4.77×10^{-6}	0.16 ± 0.29	2.29×10^{-5}
1/1320	0.3 ± 0.95	8.22×10^{-6}	0.29 ± 0.74	2.42×10^{-5}
1/1100	0.36 ± 0.95	1.33×10^{-5}	0.35 ± 0.88	1.41×10^{-5}
1/880	0.4 ± 1.15	1.52×10^{-5}	0.56 ± 1.35	3.60×10^{-5}
1/660	0.85 ± 1.65	3.85×10^{-5}	0.83 ± 1.35	5.03×10^{-5}
1/440	1.21 ± 2.28	6.34×10^{-5}	1.89 ± 2.95	1.05×10^{-4}
1/220	2.64 ± 3.28	2.18×10^{-4}	5.6 ± 6.1	3.17×10^{-4}

[a]Residual error is the vector of calibration errors at all the points.
[b]Norm of the residual error vector. It indicates goodness of fit.

Table 2. Comparison of Detection Errors of SFS and SHS

Field Bandwidth	Absolute Error (rad)		Relative Error	
	SFS[a]	SHS[a]	SFS (%)	SHS (%)
1/2200	0.002 ± 0.004	0.02 ± 0.16	1.5 ± 3	3 ± 10
1/1980	0.002 ± 0.004	0.01 ± 0.1	1.5 ± 2	2 ± 8
1/1760	0.002 ± 0.004	0.02 ± 0.15	2 ± 5	4 ± 10
1/1540	0.003 ± 0.005	0.04 ± 0.24	3 ± 7	6 ± 14
1/1320	0.004 ± 0.0075	0.01 ± 0.08	3 ± 20	3 ± 6
1/1100	0.007 ± 0.05	0.05 ± 0.3	3 ± 8	10 ± 20
1/880[b]	0.02 ± 0.27	0.1 ± 0.3	5 ± 16	12 ± 20
1/660	0.01 ± 0.05	0.1 ± 0.4	6 ± 23	15 ± 30
1/440[b]	0.04 ± 0.2	0.3 ± 0.6	9 ± 23	30 ± 50
1/220[b]	0.12 ± 0.5	0.95 ± 1.0	20 ± 110	80 ± 120

[a]$\pm$ indicates that the following quantity is the standard deviation. The error cannot be negative because it is the absolute value of the difference between the estimated phase difference and the actual phase difference.
[b]Unreasonably high standard deviation given by a possible calibration error of the phase-shift parameters; see text.

the error, the input field $f_{in}(mh, nh)$ is taken after the interpolation step to avoid including the interpolation error in the estimation error.

The SFS simulations consisted of two steps. First, we calibrated the system by finding the values of the K parameters ($\alpha_{1,2}$ included). For this we used each of the input fields as the test input. We fit the input data to the measurements using least squares with Eqs. (20) and (22). The results of the calibration are given in Table 1 for the different simulated input fields. The calibration error for K_a is under 0.05% and is not included. The results show that the error in calibrating the system increases with input bandwidth. This is due to the gradual breakdown with increasing input bandwidth of the modal representation of the output given in Eq. (3). This will make the detection error increase with input bandwidth and is considered a systematic error of the model as no noise factors were considered. Phase wrapping causes an additional increase in the estimation error. The second step was to use the calibrated parameters to estimate the amplitude and phase of the input field with Eqs. (27). We used the parameters obtained from one of the calibration tests in all the simulated measurement tests. For comparison we simulated the detection of the same input field using a SHS. The SHS consisted of an array of lenses with focal length $F = 1400$ μm and a lens period of $h = 110$ μm (the same as the SFS sampling period). The working aperture of each lens was considered to cover the entire 110 μm $\times$ 110 μm area of its corresponding period. We estimated the amplitude associated with the center of each lens or central sampling point by integrating the intensity in the SHS focal plane. We estimated the phase difference between sampling points at the opposite edges of each period (both x and y directions) in the input plane by finding the centroid of the intensity distribution in the focal plane of each lens. The intensity in the output plane is detected by an array of detectors placed on a rectangular grid

with a 1.73-μm spacing in both directions. We considered the field in the focal plane of each lens to be given by the Fourier transform of the field at the lens plane. This is justified for a slow lens ($f/10$ and slower) as described by Goodman.[19] Both the aberrations of the lenses as well as the cross talk between the output fields of adjacent lenses were neglected because this would only increase the estimation error. Neglecting the cross talk effectively amounted to a spatial frequency band limit $B = x_{max}/\lambda F \cong 0.06$ μm^{-1} for $x_{max} = 55$ μm. This is comparable to the 0.1-μm^{-1} limit in the SFS case.

The phase estimation error for both sensors, as given by the second relation in Eqs. (36), is shown in Table 2 with absolute and relative values that are normalized to the actual phase difference. The points where the error was within five standard deviations from 2π were excluded from the absolute error calculations to avoid including wrapping error. The standard deviation figure used initially was obtained by our considering only the points with an error less than 0.5 rad. Also, the points where the input phase difference was smaller than 10^{-2} rad were additionally excluded from the relative error calculations. Note that 97% on average and no less than 93% of the total number of points were included in the calculations in all cases for both sensors. The amplitude error is not shown as it is practically zero for the SFS and ranges from 5–100% for the bandwidths considered in the SHS case. Note that the standard deviations for the 1/880-, 1/440-, and 1/220-μm^{-1} entries in Table 2 are much bigger than in the rest of the cases and are comparable in size to π. This is because there are a few points in the input phase differences that are close to π or $-\pi$ and are estimated to be at the opposite side, i.e., close $-\pi$ and π, respectively. The possible cause for this is an error in calibrating the $\alpha_{1,2}$ parameters corroborated with the wrapping of the phase. As explained above, this error increases with input field bandwidth (see
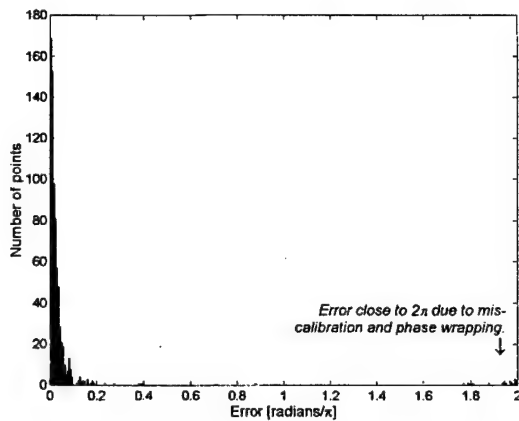
Fig. 4. Histogram of phase estimation error at maximum input bandwidth ($1/220$ μm^{-1}). The size of the SFS sampling grid is 20×20, thus there are 760 phase estimation points. There are 14 points with an error within 2 rad of 2π.
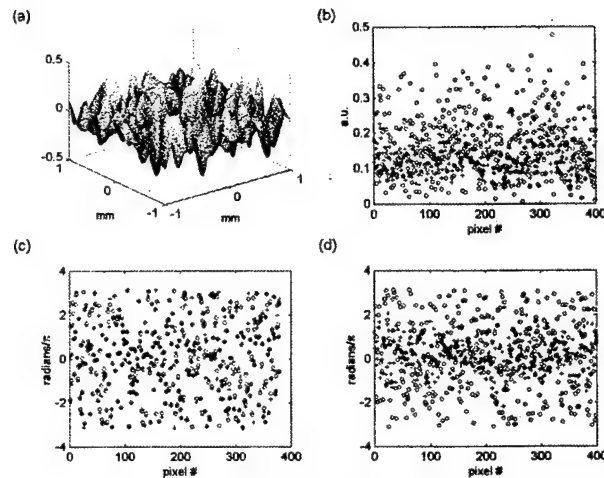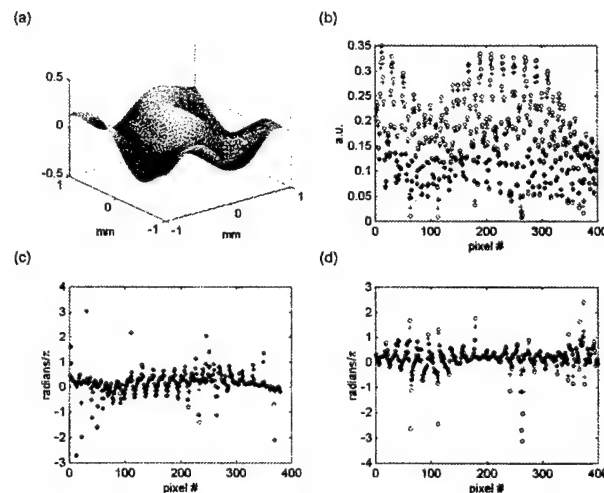

Fig. 5. Detection of a random input field with a spatial bandwidth of $1/220$ μm^{-1}. The input values are marked with circles and the estimated values are marked with crosses: (a) input field (real part) with $B = 1/220$ μm^{-1}, (b) SHS amplitude detection, (c) SFS phase detection, and (d) SHS phase detection.

Table 1). A possible remedy would be to use more than two phase pixels having different phase shifts. The inversion algorithm would then use this redundant information to reduce the error. If we eliminate the points for which the difference between the input and the detected phase difference is within 2.0 rad of 2π, the errors change to 0.01 ± 0.02 rad for the $1/880$-μm^{-1} case, to 0.03 ± 0.05 rad for the $1/440$-μm^{-1} case, and to 0.1 ± 0.15 rad for the $1/220$-μm^{-1} case. Note that only 2 points in the first case, 4 points in the second, and 14 points in the third case had to be eliminated. This is from a total of 760 phase differences. Another way to see that there are just few points around $\pm\pi$ that cause this problem is to plot the histogram of the error vector (difference between the input and the detected values) as we show in Fig. 4 for the $1/220$-μm^{-1} case.

Although limited in scope, the simulations show a considerably smaller error for the SFS method as opposed to the SHS method. In a real situation, however, the poor light throughput of this particular SFS implementation is likely to tip the balance of the comparison in the opposite direction. To quantify this statement we consider the detection SNR as limited by the shot noise only. We consider SNR_{SFS} and SNR_{SHS} to be the signal-to-noise ratios of the two systems. For the case described here, the ratio SNR_{SFS}/SNR_{SHS} will be approximately equal to the ratio of the power throughput of the two systems, which is of the order of 10^{-2}. This means that 10^2 more power is required by the SFS to have the same shot-noise-limited SNR as the SHS. This implementation of the fan-out stage is a simple one meant for testing the sensor concept. Other implementations, from the realm of array generation techniques, can potentially increase the power throughput by 2 orders of magnitude. On the other hand, its simplicity and real-time phase detection capability may make it attractive for applications such as high-energy laser testing. We also note that our sensor concept has a $(-\pi, \pi)$ dynamic range, similar to any

shearing interferometer. The SHS does not have this limitation.

Figures 5 and 6 show the results obtained with both sensors for two types of input fields. Figure 5 shows the maximum input bandwidth and Figure 6 shows a typical low-bandwidth field. The amplitude detection error is shown only for the SHS because it is small in the SFS case. Each set of figures contains a view of the real part of the input field, the estimated amplitude versus the input amplitude, and the estimated phase difference versus the input phase difference. The estimated values are marked in the figures with crosses and the input values with circles for each of the 400 detection points. Perfect estima-

Fig. 6. Detection of a random input field with a spatial bandwidth of $1/2200$ μm^{-1}. The input values are marked with circles and the estimated values are marked with crosses: (a) input field (real part) with $B = 1/2200$ μm^{-1}, (b) SHS amplitude detection, (c) SFS phase detection, and (d) SHS phase detection.
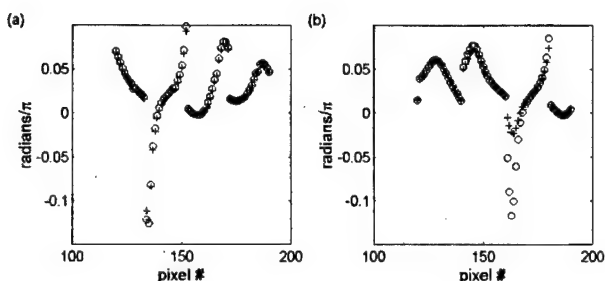
Fig. 7. SHS and SFS phase detection error for low input bandwidth. The input values are marked with circles and the estimated values are marked with crosses: (a) SFS detection and (b) SHS detection.

tion occurs every time the cross is centered in the circle for the respective detection point.

The poor amplitude and phase estimation of the SHS are due to the fact that the values of the field amplitude and phase at the sampling point are close to the values estimated only when the input field has a low enough spatial bandwidth. This is because the sensor estimates the average values of the two quantities[10] for each subaperture. According to the mean value theorem, these values are close to the actual values at the sampling points only for small enough input bandwidths. This explains why the output of the SHS for $1/2200$ μm^{-1} is much better than for $1/220$ μm^{-1}. Also, the SHS still shows considerable error compared with the SFS at certain estimation points, even for low input bandwidths. Figure 7 shows a detail of the graphs in Figs. 6(c) and 6(d) to illustrate this fact.

B. Experimental Results

We performed a set of experiments to further test our method. Ideally we want to find the K parameters to be able to estimate an arbitrary field. Here we propose a first step toward that goal: input a tilted plane wave to the system and check if the intensities in the phase pixels show a sinusoidal variation with the change in incidence angle. The experimental setup is shown in Fig. 8. It consists of two cascaded $4f$ systems that form the image of a sampling mask on a CCD camera. The first $4f$ system had a 3-mm hard-edge low-pass aperture in the Fourier plane and a magnification of 2.0×. The focal lengths of the two lenses were 50 and 100 mm, respectively. The second cascaded $4f$ system was identical to the first one but without any stop in its Fourier plane. The CCD pixel size was 22 × 22 μm. As discussed in Subsection 2.B, a small enough receiver pixel size is necessary to obtain good modulation in the phase pixels. The overall magnification of 4.0× ensured an effective receiver pixel size of around 5 μm, close to the value in the simulations. The sampling mask consisted of 10-μm-square holes spaced at 110 μm. It was placed approximately 1 mm in front of the object plane of the system. The input to the system was a collimated beam coming from a laser diode and steered with a mirror mounted on a rotation stage.
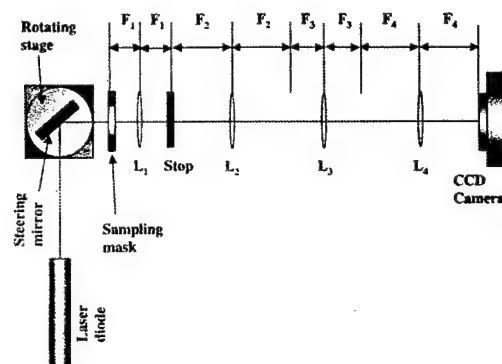
Fig. 8. Experimental setup. Sampling mask is a 40 × 40 array of 10-μm-square holes in a chrome coating over a 1-mm-thick quartz substrate. The focal lengths of the lenses L_1, L_2, L_3, and L_4 are $F_1 = F_2 = 50$ mm and $F_3 = F_4 = 100$ mm. The sampling mask is placed approximately 1 mm in front of the front focal plane of L_1. The aperture of the stop in the Fourier plane of the first lens is 3 mm. The camera is a 576 × 384 CCD array with 22-μm-square pixels.

The wavelength of the laser radiation was between 620 and 680 nm according to manufacturer's specifications. Our measurements indicated a value of 690 ± 20 nm. The output of the system at normal incidence is shown in Fig. 9(d). The log base 10 of the measured intensity is shown. The experiment consisted of rotating the steering mirror over a $[-0.2°, +0.2°]$ interval to scan the incidence angle. The signal at the amplitude and the phase pixels was recorded. For a clean measurement, the tilted plane wave model of the experiment predicts a sinusoidal modulation in the phase pixels of a period given by

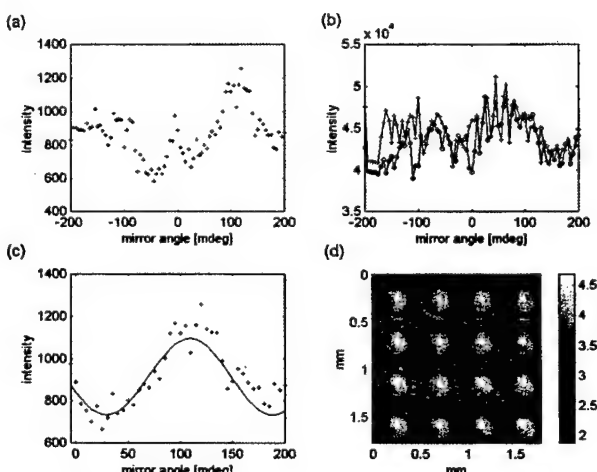$$T = \frac{\lambda}{2h} \frac{180}{\pi} 1000 \text{ mdeg.} \qquad (37)$$

Fig. 9. Experimental results: (a) signal in a phase pixel as a function of mirror tilt angle, (b) signal in two adjacent amplitude pixels as a function of mirror tilt angle, (c) signal in the same phase pixel for a smaller range of angles (working bandwidth) with a sinusoid fitting it, and (d) intensity (log 10) in the output plane of the device.
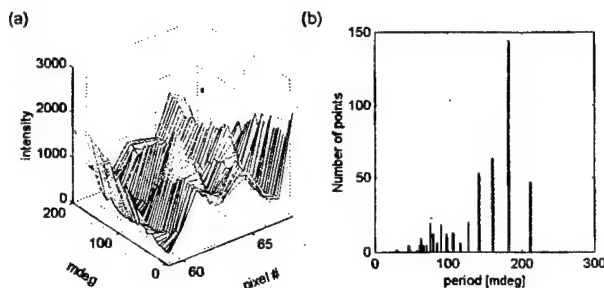
Fig. 10. Experimental verification. (a) Change in the receiver pattern function with angle of incidence. The receiver pattern function is different for different values of the incidence angle. This is due to the failure of the SFS representation to model the output field at high input bandwidths (the model breaks down). (b) Histogram of the measured period for 300 phase pixels. The average period is approximately 150 mdeg with a standard deviation of 40 mdeg which is in agreement with the theoretical estimate.

Assuming $\lambda = 680$ nm, Eq. (34) yields a value $T = 177$ mdeg of stage rotation angle. Figure 9(a) shows the intensity at a phase pixel [pixel (63,63) in Fig. 9(d), which is 81×81], and Fig. 9(b) shows the intensity in the corresponding x adjacent amplitude pixels. We note the modulation of the signal in the amplitude pixels that is due to variable intensity in the input wave front. This degrades the ideally sinusoidal signal in the phase pixels. The phase signal has low-frequency components because of a number of other factors that are not related to the phase difference between the two adjacent sampling points. The most important factor is the change in the receiver pattern function with incidence angle. As pointed out by the numerical simulations, the representation in Eq. (3) is valid on a limited range of input spatial frequencies. To measure the sinusoidal modulation period we scanned the incidence angle and we exceeded the angular range where Eq. (3) is valid. Therefore we could not use it to represent the detected signal over the entire range of angles. In fact, the receiver pattern functions shift laterally and, when the incidence angle is increased over a certain value, the shift cannot be neglected. The fact that we use a hard-edge stop in the Fourier plane as opposed to the windowed Fourier transform used in the simulations makes the receiver pattern functions even more dependent on the incidence angle of the input field because of the more pronounced sidelobes. Figure 10(a) shows the change in the receiver patterns with incidence angle. Restricting the data to a 200-mdeg range roughly corresponding to the $[-1/2h, 1/2h]$ working bandwidth, we obtain a modulation period of approximately 160 mdeg, close to the one predicted by Eq. (34). Figure 9(c) shows the signal in this range as well as the fitted sinusoidal modulation. In fact, we considered the signals from 300 phase pixels placed in the middle region between adjacent amplitude pixels. The amplitude pixels were placed at the maxima of the intensity patterns. Figure 10(b) shows the histogram of the measured

periodicity distribution. The average value is approximately 150 mdeg with a standard deviation of 40 mdeg, which is in agreement with the theoretical estimation from Eq. (34). Also, our experiments emphasized the real-time capability of the SFS. We attempted to make the same measurement but using a phase-stepped Mach–Zehnder shearing interferometer instead. We did not obtain meaningful data until we decreased the shear to approximately 16 μm as opposed to 110 μm in the SFS case. This was due to the beam pointing error of the laser diode and the fact that we were using a multiframe technique. We used the measurements taken with the Mach–Zehnder interferometer to estimate a sinusoidal period of 182 ± 10 mdeg, confirming our previous estimate.

C. Discussion

We tested the consistency of the SFS method both numerically and experimentally. Numerical simulations proved the wave-front sensing principle of this device. We showed experimentally that the signal detected in the phase pixel is proportional to the cosine of the phase difference between two adjacent sampling points. The most important drawback of the current implementation of the fan-out stage is the low light throughput. Alternative methods, such as array generation techniques, may also be considered. For example, one can use diffractive masks placed in the Fourier plane of the $4f$ system. This could increase the light throughput by 2 orders of magnitude. Also, one can use matrices of diffractive lenses instead of the refractive lenses used in this research. By reducing the focal lengths, we could make the system more compact, possibly by 2 orders of magnitude. Ideally, use of an array generator as the fan-out element would create multiple phase-shifted replicas of the sampling mask at the receiver plane, thus making possible use of one-shot phase-stepping interferometry techniques. Use of the sampling stage is the key element to allow for allocating the blocked regions in the input to measurement regions in the output. However, the setup that we presented has the advantage of being easy to implement. Equations (13) suggest that we can eliminate the $4f$ system if we could make the sampling function $s_0(x)$ band limited, like a tapered hole, for example. In this way the first two factors of the convolution could be replaced by a tapered sampling hole without changing the overall operation of the system. The tapered-hole SFS sampling mask would be placed approximately 1 mm in front of a CCD array. This simple improvement would make the system considerably smaller and thus less sensitive to vibrations. Materials for true gray-level masks are commercially available.

One problem that must be solved, before a successful implementation of the SFS can be accomplished, is the calibration procedure. Obtaining the values of the K parameters was easy in the simulation because we had access to the input field directly. This was not the case for the experiment. Having greater

light throughput as well as a better SNR of the signal in the phase pixels would allow use of more-refined calibration procedures.

4. Conclusions

We proposed a new wave-front sensing method based on reconstructing the input field from its samples and demonstrated a particular sensor implementation. Numerical tests showed that it can detect the phase of input fields up to the Nyquist limit with an error increasing with bandwidth but lower than 20%. Experimental tests showed the consistency of the model by correctly measuring the tilt of a plane wave at a variable incidence angle. However, we were unable to solve the system calibration problem and therefore reconstruct arbitrary fields. Alternative techniques similar to array generation may be used in designing the fan-out stage of the device and improving its light efficiency. The fact that the SFS could sense fields with bandwidths up to the Nyquist limit make it a good candidate for applications that require detection of high information content fields. Its one-shot phase detection capability could benefit a number of applications needing real-time full-field detection. In summary, the main qualities of the SFS are its potential compactness, ease of use, and real-time phase detection. Compactness reduces its sensitivity to vibrations and, together with the other two characteristics, could help spread use of optical phase detection systems outside the laboratory environment.

References

1. M. C. Roggemann and B. Welsh, *Imaging through Turbulence* (CRC Press, Boca Raton, Fla., 1996).
2. S. A. Klein, "Optimal corneal ablation for eyes with arbitrary Hartmann–Shack aberrations," J. Opt. Soc. Am. A **15**, 2580–2588 (1998).
3. J. Pfund, N. Lindlein, J. Schwider, R. Burow, T. Blumel, and K. E. Elssner, "Absolute sphericity measurement: a comparative study of the use of interferometry and a Shack–Hartmann sensor," Opt. Lett. **23**, 742–744 (1998).
4. M. Zajac and B. Dubik, "Measurement of wavefront aberrations of diffractive imaging elements," in *Tenth Polish-Czech-Slovak Optical Conference: Wave and Quantum Aspects of Contemporary Optics*, J. Nowak and M. Zajac, eds., Proc. SPIE **3320**, 237–241 (1998).
5. M. C. Roggemann, B. M. Welsh, and R. Q. Fugate, "Improving the resolution of ground-based telescopes," Rev. Mod. Phys. **69**, 437–505 (1997).
6. M. C. Roggemann, B. M. Welsh, P. J. Gardner, R. L. Johnson, and B. L. Pedersen, "Sensing three-dimensional index-of-refraction variations by means of optical wavefront sensor measurements and tomographic reconstruction," Opt. Eng. **34**, 1374–1384 (1995).
7. J. M. Geary, *Introduction to Wavefront Sensors*, Vol. TT18 of SPIE Tutorial Text (SPIE Press, Bellingham, Wash., 1995).
8. R. N. Smartt and W. H. Steel, "Theory and application of point-diffraction interferometers (telescope testing)," Jpn. J. Appl. Phys. **14**, 351–356 (1975).
9. Y. Baharav, B. Spektor, J. Shamir, D. G. Crowe, W. Rhodes, and R. Stroud, "Wave-front sensing by pseudo-phase-conjugate interferometry," Appl. Opt. **34**, 108–113 (1995).
10. R. K. Tyson, *Principles of Adaptive Optics* (Academic, Boston, 1991).
11. F. Roddier, "Curvature sensing and compensation: a new concept in adaptive optics," Appl. Opt. **27**, 1223–1225 (1988).
12. D. L. Fried, "Least-square fitting a wave-front distortion estimate to an array of phase-difference measurements," J. Opt. Soc. Am. **67**, 370–375 (1977).
13. R. J. Noll, "Phase estimates from slope-type wave-front sensors," J. Opt. Soc. Am. **68**, 139–140 (1978).
14. W. H. Southwell, "Wave-front estimation from wave-front slope measurements," J. Opt. Soc. Am. **70**, 998–1006 (1980).
15. R. C. Cannon, "Global wave-front reconstruction using Shack-Hartmann sensors," J. Opt. Soc. Am. A **12**, 2031–2039 (1995).
16. R. Cubalchini, "Modal wave-front estimation from phase derivative measurements," J. Opt. Soc. Am. **69**, 972–977 (1979).
17. J. Herrmann, "Cross coupling and aliasing in modal wave-front estimation," J. Opt. Soc. Am. **71**, 989–992 (1981).
18. G. Toraldo di Francia, "Degrees of freedom of an image," J. Opt. Soc. Am. **59**, 799–804 (1969).
19. J. W. Goodman, *Introduction to Fourier Optics*, 2nd ed. (McGraw-Hill, New York, 1996).
20. D. W. Robinson and G. T. Reid, *Interferogram Analysis* (Institute of Physics Publishing, Philadelphia, Pa., 1993).
21. J. J. Knab, "Interpolation of band-limited functions using the approximate prolate series," IEEE Trans. Inf. Theory **IT-25**, 717–720 (1979).

Rotational Shear Interferometers

Visible Cone-Beam Tomography With a Lensless Interferometric Camera

Daniel L. Marks,[1,2] Ronald A. Stack,[1] David J. Brady,[1,2]*
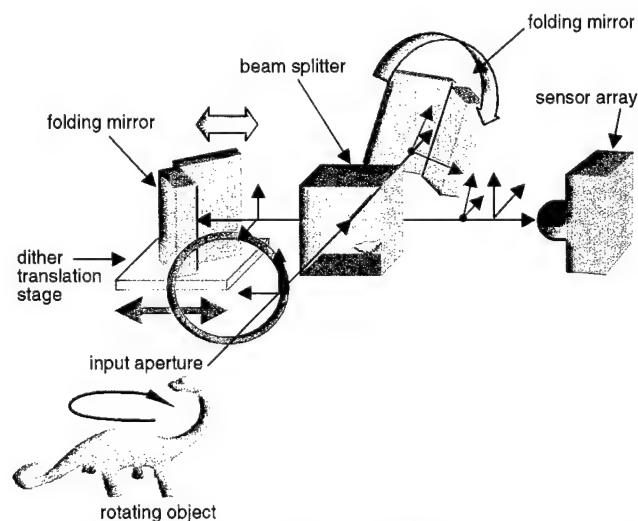David C. Munson Jr.,[1,2] Rachael B. Brady[3]

Digital processing of optical coherence functions can reconstruct three-dimensional objects illuminated by incoherent light. It is shown that Fourier analysis of the mutual intensity of the field produces projections that are mathematically identical to the projections of x-ray cone-beam tomography. A lensless interferometric camera that captures planes of mutual intensity data is described and used to reconstruct an incoherently illuminated visible object in three dimensions.

Lenses act as analog computers that transform the incident field into an image of the field in a particular plane. With the continuing digital revolution, one may wonder whether this analog processing can be digitally enhanced. This report describes digital imaging with an optical system consisting only of smooth planar surfaces. As is often the case when a digital processor replaces an analog one, our motivation is to improve the analog algorithm. The improvement we obtain is infinite depth of focus, which is equivalent to the geometrical optics assumption that the field propagates in nondiffracting rays. This assumption is satisfactory in medical x-ray tomography because one is satisfied to resolve features that are large compared with the wavelength of the illuminating radiation. With visible imaging, one often wishes to resolve features as close to the wavelength scale as possible, in which case diffraction cannot be neglected. Here we show that visible ray projection data obtained from digital analysis of interferometric data can be combined with tomographic algorithms to reconstruct three-dimensional (3D) objects. Our results show that neither point-by-point scanning, as in confocal microscopy or coherence tomography, nor heuristic analysis, as in computer vision, is necessary for 3D reconstruction and that diffraction-limited 3D optical reconstruction is possible from purely physical field analysis.

We obtained infinite depth of focus images by digital analysis of the mutual intensity function. For quasi-monochromatic light, the mutual intensity between two points is $J_{12} = \langle E_1 E_2^* \rangle$, where $\langle \ \rangle$ is the statistical expected value. E_1 and E_2 are the complex field values at points 1 and 2, and we considered scalar fields for sim-

plicity (1). If the field arises from a primary source in free space, its value at point 1 is a superposition of Huygens wavelets. This superposition can be expressed as $E_1 = \int E_s(e^{jkR_{1s}}/R_{1s})d^3r_s$, where j represents $\sqrt{-1}$, E_s is the source field density, $k = 2\pi/\lambda$, and position vector r_s is the variable of integration. The integral is over the source volume, and R_{1s} is the distance from a source point to point 1. For a spatially incoherent source, $\langle E_s E_{s'}^* \rangle = I_s \delta(r_s - r_{s'})$, where I_s is the source intensity density due to the field fluctuations E_s and $E_{s'}$ at points r_s and $r_{s'}$, respectively, and $\delta(\ )$ is the Dirac delta function. After double integrations over r_s and $r_{s'}$, the expectation reduces to $J_{12} = \int I_s(e^{jk(R_{1s}-R_{2s})}/R_{1s}R_{2s})d^3r_s$. Interferometric astronomical imaging uses a far-field approximation of this integral in which the source space reduces to 2D and the exponential term becomes a Fourier transform kernel (2, 3). The integral can be inverted to obtain the 3D source density with Fourier or modal methods (4–8). High depth of focus has been studied in the context of statistical radiometry (9).

To use the mutual intensity to obtain cone-beam projections, we assumed that source point r_s was confined to a semiinfinite region $z_s > 0$ with Cartesian coordinates (x_s, y_s, z_s) and the coherence sampling points (for example, points 1 and 2) were confined to a planar aperture on $z_s = 0$. We chose the origin of the aperture at the midpoint between the two sampling points and defined Δx and Δy to be the distance between the sampling points along the x and y axes. The Cartesian coordinates of the sampling points are ($\Delta x/2$, $\Delta y/2$,0) and ($-\Delta x/2$, $-\Delta y/2$,0). Finally, we made the paraxial approximation that $z_s \gg \Delta x, \Delta y, x_s, y_s$ for all points in the source volume and in the correlation aperture. This implies that $R_{1s} \approx z_s + [(\Delta x/2 - x_s)^2 + (\Delta y/2 - y_s)^2]/2z_s$ and $R_{2s} \approx z_s + [(\Delta x/2 + x_s)^2 + (\Delta y/2 + y_s)^2]/2z_s$. Under this approximation,

$$J(\Delta x, \Delta y) =$$

$$\int \frac{I_s}{z_s^2} \exp\left[-\frac{j2\pi}{\lambda z_s} \left(x_s\Delta x + y_s\Delta y \right) \right] d^3r_s \quad (1)$$

where λ is the center wavelength of the source. Taking the inverse Fourier transform
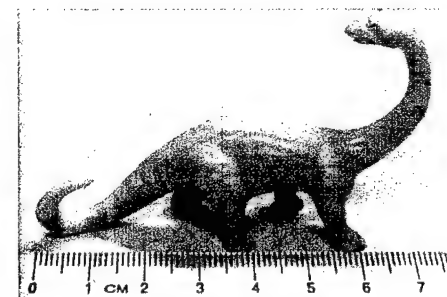
Fig. 2. Photograph of the test object. The maximal length of the object is 7.2 cm, the width is 2.1 cm, and the height is 4.9 cm.

Fig. 1. The rotational shear interferometer (RSI) is a two-arm Michelson-style interferometer. A folding mirror consisting of a pair of planar mirrors joined at right angles terminates each arm. Each folding mirror inverts the incident field across its axis. The RSI measures planes of interference data in parallel with an electronic sensor array in the output aperture. The folding mirrors are nominally placed so that the optical path difference between the arms is zero. The interference is separated from background terms by dithering the relative optical path delay with a translation stage on one arm. The coordinate system of the object corresponds to (x_s, y_s, z_s), and the plane of the sensor array corresponds to the correlation space (Δx, Δy).

[1]Beckman Institute for Advanced Science and Technology, [2]Department of Electrical and Computer Engineering, [3]National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, 405 North Mathews Avenue, Urbana, IL 61801, USA.

*To whom correspondence should be addressed. E-mail: dbrady@uiuc.edu

of Eq. 1 with respect to Δx and Δy, we obtained

$$\tilde{J}(u,v) =$$

$$\int J(\Delta x, \Delta y) \exp[j2\pi(u\Delta x + v\Delta y)] \, d\Delta x d\Delta y$$

$$= \int \frac{I_s}{z_s^2} \delta\left(u - \frac{x_s}{\lambda z_s}, v - \frac{y_s}{\lambda z_s}\right) d^3 r_s$$

(2)

$\tilde{J}(u,v)$ is a line integral through I_s/z_s^2 along a ray passing through the points ($x_s = \lambda z_s u$, $y_s = \lambda z_s v$, z_s). Values of $\tilde{J}(u,v)$ for all allowed values of u and v correspond to integrals along a cone of rays diverging from the vertex point ($x_s = y_s = z_s = 0$). In x-ray tomography, a cone of projection data is gathered by placing a planar sensor on the opposite side of the object volume from a point source. Equation 2 shows that a mathematically equivalent cone of data for a self-luminous or ambiently illuminated visible object is obtained by measuring the mutual intensity on a plane centered on the equivalent (but now virtual) point source.

Planes of mutual intensity data may be measured in parallel with a rotational shear interferometer (RSI) (Fig. 1) (10–13). We obtained experimental data using an RSI formed of a 5-cm aperture beam splitter and 5-cm folding mirrors. One of the folding mirrors was mounted on a piezo-driven flexture stage to vary the optical path length. The only other elements in the optical system were a mechanical shutter at the RSI input, a 3-nm bandpass spectral filter centered on a wavelength of 633 nm at the output plane, and a 512 pixel by 512 pixel back-illuminated charge-coupled device (CCD) detector array. The spectral filter enforces the quasi-monochromatic assumption. For a quasi-monochromatic field, an RSI isolates the amplitude and phase of the mutual intensity by sampling the output plane intensity as a function of optical path difference. We measured the output for eight optical path delays between the two

arms. The eight delays are evenly spaced over one wavelength of maximal relative delay. The discrete Fourier transform of the intensity image over these eight frames is $J(\Delta x, \Delta y)$.

Cone-beam tomography uses ray projections through vertices lying on a curve called the vertex path. Exact reconstruction of an object volume is possible if all planes through the object volume intersect the vertex path (14). Vertex paths that sample incomplete data are often used for implementation simplicity. We used an algorithm from Feldkamp *et al.* (15) in our experiments. This algorithm is based on a circular vertex path. Our test object (Fig. 2) was placed 1.61 m from the RSI sensor plane and illuminated by a white halogen lamp. We sampled a circular vertex path by rotating the object in front of the RSI. Planes of coherence data were recorded from 128 vertex points equally spaced in angle over one revolution. At each vertex point, we captured eight frames of 128 by 128 intensity samples. These frames were demodulated to estimate $J(\Delta x, \Delta y)$, which was then Fourier transformed to obtain 128 planes of $\tilde{J}(u,v)$ data. These planes were used in the cone-beam algorithm to reconstruct the 128 by 128 by 128 data volume (Fig. 3). The reconstructed data cube is 10.6 cm on a side with a resolution of 830 μm. The object size and resolution are determined by the range and sampling rate of ($\Delta x, \Delta y$). The sampling rate was the CCD pixel spacing (22 μm), and the range was limited by the RSI aperture (limited by the CCD array size to 0.63 cm) and by the angle between the fold axes (6.55°) to 0.7 mm (16).

Our derivation assumes that the object is translucent, but our experiment reconstructs an opaque object. Opacity has surprisingly little effect for objects without occluding surfaces. The tomographic reconstruction of a convex opaque object is a linear superposition of reconstructions of the differential surface patches that make up the object. The opacity of each surface patch can be modeled as a window on the solid angle over which the patch radiates.

The window function produces a characteristic patch response oriented according to the patch's surface normal (Fig. 4). For a convex object, the reconstruction is the convolution of the surface with the patch response function. In the nonconvex case, surface patches may obscure each other, resulting in the reconstruction no longer being a unique function of the surface structure. The volume surrounding the feet of the object reported here is nonconvex, which leads to uncertainty in the reconstruction of this region (Fig. 5).

Systems combining digital computation with a coherence sensor such as the RSI can achieve infinite depth of field. This property makes cone-beam tomography a flexible tool to synthesize 3D structure from coherence information. Such physical optics techniques may ultimately benefit microscopy and machine vision by providing 3D reconstructions of supe-
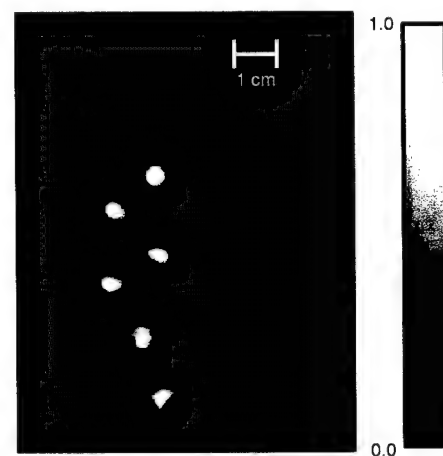

Fig. 5. Slice $z = 84$ of the data volume showing the legs, the tail going down, and the tail coming up. The four top white circles are cross sections of the legs and the bottom two circles are cross sections of the tail. The "fill" between the legs and around the tail is due to the angular windowing resulting from occlusion of some patches.

Fig. 3 (left). A pseudocolor volume rendering of the 128 by 128 by 128 reconstructed data volume. The haze around the reconstructed data volume in Fig. 3 is due to the spatial distribution of the patch response function. **Fig. 4 (right).** The data volume with planes slicing the neck and body of the dinosaur. The brightness on the planes corresponds to the reconstructed intensity density. The spatial distribution of the patch response function is readily visible on the body slice.
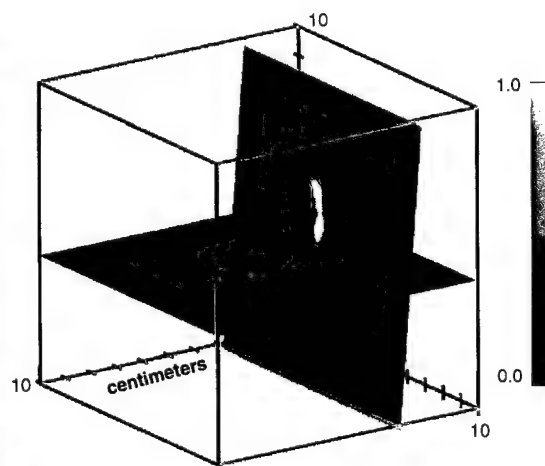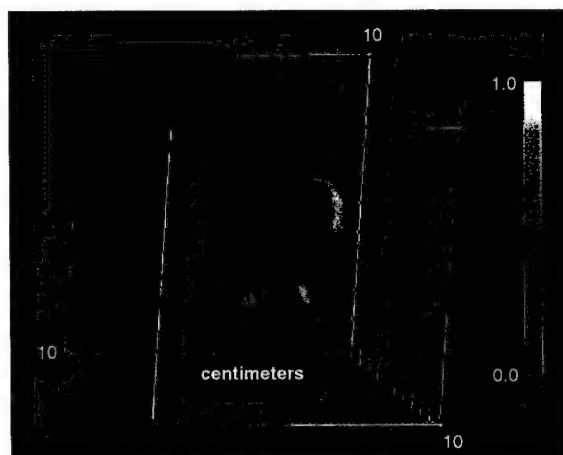
rior resolution. Nonimaging sensors may provide advantages over lens-based cameras, because our knowledge of the environment should be limited by the information available from it and not our sensing or computational methods, analog or digital.

References and Notes

1. L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics* (Cambridge Univ. Press, Cambridge, 1995).
2. C. V. Schooneveld, Ed., *Image Formation from Coherence Functions in Astronomy, IAU Colloquium no. 49* (Reidel, Groningen, Netherlands, 1978), vol. 76.
3. G. W. Swenson, *J. Opt. Soc. Am. A* **3**, 1311 (1986).
4. W. H. Carter and E. Wolf, *Opt. Acta* **28**, 227 (1981).
5. A. J. Devaney, *J. Math. Phys.* **20**, 1687 (1979).
6. I. J. LaHaie, *J. Opt. Soc. Am. A* **2**, 35 (1985).
7. J. Rosen and A. Yariv, *Opt. Lett.* **21**, 1803 (1996).
8. A. M. Zarubin, *Opt. Commun.* **100**, 491 (1993).
9. K. Yoshimori *et al.*, *J. Opt. Soc. Am. A* **14**, 3379 (1997).
10. J. D. Armitage and A. Lohmann, *Opt. Acta* **12**, 185 (1965).
11. F. Roddier, paper presented at the Proceedings of the International Astronomical Union Colloquium, Sydney NSW, Australia, 1979.
12. K. Itoh and Y. Ohtsuka, *J. Opt. Soc. Am. A* **3**, 94 (1986).
13. K. Itoh *et al.*, *Jpn. J. Appl. Phys.* **29**, L1561 (1990).
14. H. K. Tuy, *SIAM J. Appl. Math.* **43**, 546 (1983).
15. L. A. Feldkamp *et al.*, *J. Opt. Soc. Am. A* **1**, 612 (1984).
16. D. L. Marks *et al.*, *Appl. Opt.* **38**, 1332 (1999).
17. Supported by the Defense Advanced Research Projects Agency. D.L.M. acknowledges the support of an NSF Graduate Fellowship.

Chiral Magnetic Domain Structures in Ultrathin FePd Films

H. A. Dürr,[1] E. Dudzik,[1,2] S. S. Dhesi,[1] J. B. Goedkoop,[3] G. van der Laan,[1]* M. Belakhovsky,[4] C. Mocuta,[4] A. Marty,[4] Y. Samson[4]

The magnetization profile of magnetically ordered patterns in ultrathin films was determined by circular dichroism in x-ray resonant magnetic scattering (CDXRMS). When this technique was applied to single crystalline iron palladium alloy layers, magnetic flux closure domains were found whose thickness can constitute a large fraction (~25 percent) of the total film.

X-ray reflections only occur when equivalent sites in a crystal are occupied by identical atoms. If the scattering amplitudes of equivalent sites are not the same, then forbidden reflections can occur. These are pronounced in the case of resonant diffraction, where virtual excitations from core to valence states impose the symmetry properties of the electronic and magnetic structure of the material (*1*). For instance, an antiferromagnetic ordering will give a magnetic superlattice with twice the size of the charge distribution. Here, we show how resonant magnetic scattering can be used to study complicated closure domain patterns (Fig. 1).

The domains display a left-right handedness known as chirality. It can be verified that the magnetization direction of each of the bulk domains in Fig. 1 is related to the magnetization of the closure domains right (left) above by a (counter)clockwise quarter-turn rotation in the *yz* plane. This extra symmetry condition should correspond to an additional Bragg condition, leading to an otherwise forbidden reflection. Although the possibility of measuring the long-period magnetic structure

by magnetic x-ray scattering was suggested by Blume in 1985 (*2*) and has been successfully applied to magnetic lattice periodicities on an atomic scale (*3*), we demonstrate here the case of magnetic domain structures. Using x-rays with circular polarization, we can make an unambiguous distinction between magnetic profiles with ↑ → ↓ ← ↑ and ↑ ↓ ↑ ↓ domain patterns because only the former has a chiral structure. The observation of circular dichroism in the x-ray resonant magnetic scattering (CDXRMS) signal, *I*—that is, its difference between left and right circularly polarized photons—allows us to recover the phase information that is generally lost in diffraction experiments. We demonstrate that this effect can be directly related to the magnetization profile in the film.

To observe the magnetization directions, we can use the equivalent in the x-ray region of either the Faraday rotation of linearly polarized light or the Kerr effect of elliptically polarized light. An increase in the sensitivity for the valence electron magnetization is obtained by tuning the photon energy to the Fe L_3 edge (wavelength $\lambda = 17.5$ Å), where a $2p$ core electron is excited into an empty, magnetically aligned $3d$ state. This wavelength is of the correct magnitude to be susceptible to the magnetic periodicity of the sample. The scattering signal measured in a diffraction experiment, $I \propto |\Sigma_n \exp(i\mathbf{q}\cdot\mathbf{r}_n)f_n|^2$ (where **q** is the photon wave vector transferred in the scattering process), is the square of the modulus of the sum over all lattice sites, $\mathbf{r}_n$, of

the scattering amplitudes, f_n, weighted by a phase factor (*4*). Hannon *et al.* (*5*) showed that the resonant electrical dipole scattering amplitude can be written as

$$f_n = \hat{\mathbf{e}}' \cdot \hat{\mathbf{e}}\, F_n^{(0)} - i(\hat{\mathbf{e}}' \times \hat{\mathbf{e}}) \cdot \hat{\mathbf{M}}_n\, F_n^{(1)}$$
$$+ (\hat{\mathbf{e}}' \cdot \hat{\mathbf{M}}_n)(\hat{\mathbf{e}} \cdot \hat{\mathbf{M}}_n)\, F_n^{(2)} \qquad (1)$$

where $\hat{\mathbf{e}}$ and $\hat{\mathbf{e}}'$ are the polarization vectors of the incident and scattered x-rays, respectively, and $\hat{\mathbf{M}}_n$ is the unit vector along the magnetization direction in the sample. The complex factors F_n describe the atomic resonant excitation and decay processes, and they can be expanded in terms of multipole moments of the ground state (*6*). The first term in Eq. 1 is due to scattering from the Fe charge distribution, whereas the latter two terms are purely magnetic scattering contributions. In the following we use the second term in Eq. 1 to reconstruct the magnetization profile of the film. The difficulty with this is that usually the absolute magnitude of the complex factors F_n is not very well known and can only be obtained directly under certain conditions, such as for multilayered samples (*7, 8*). However, the case of regular domain patterns results in an elegant way to separate the three scattering contributions in Eq. 1. The lateral domain periodicity leads to purely magnetic superstructure scattering peaks located symmetrically around the specularly reflected x-ray beam. For structurally well-ordered films with smooth interfaces, the charge scattering term in Eq. 1 contributes only to the specular peak. The two magnetic terms are linear and quadratic in $\hat{\mathbf{M}}_n$ and cause magnetic peaks at wave vectors $\pm\tau$ and $\pm 2\tau$, respectively ($2\pi/\tau$ is the domain periodicity) (*3–5*).

To assess the scattering from the individual domains in Fig. 1B, we must determine the scattering cross sections for the x-ray polarization components σ and π that are perpendicular and parallel to the scattering plane, respectively (*9*). For the scattering geometry used (Fig. 2A) and concentrating on the second term in Eq. 1, there are mainly two scattering paths producing π-polarized scattered light (*4*). For the bulk domains, $\hat{\mathbf{M}}_n$ is perpendicular to the film and σ-polarized incident radiation experiences a Faraday rota-

[1]Daresbury Laboratory, Magnetic Spectroscopy Group, Warrington WA4 4AD, UK. [2]University of York, York YO1 5DD, UK. [3]University of Amsterdam, Valckenierstraat 65, NL 1018 XE Amsterdam, Netherlands. [4]CEA/Grenoble, Service de Physique des Matériaux et Microstructures, 17 rue des Martyrs, 38054 Grenoble Cedex 9, France.

*To whom correspondence should be addressed. E-mail: g.vanderlaan@dl.ac.uk

SPATIO-SPECTRAL TRIANGULATION OF VISIBLE AND INFRARED POINT SOURCES USING A PORTABLE ROTATIONAL SHEAR INTERFEROMETER

BY

JASON RICHARD GALLICCHIO

B.S., University of Illinois at Urbana-Champaign, 1999

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2001

Urbana, Illinois

SPATIO-SPECTRAL TRIANGULATION OF VISIBLE AND INFRARED POINT SOURCES USING A PORTABLE ROTATIONAL SHEAR INTERFEROMETER

Jason Richard Gallicchio, MS
Department of Electrical Engineering
University of Illinois at Urbana-Champaign, 2001
David Brady, Adviser

I show how a rotational shearing interferometer (RSI) can determine the location and spectral radiance of a sparse array of point sources. The approach is particularly applicable to situations in which one seeks to estimate a relatively compact set of target positions and spectral components, rather than a pointwise spatio-spectral data cube. In these situations, RSI triangulation provides efficient aperture limited source localization and spectral analysis. I describe a mobile version of an RSI module with built-in electronics and a computer with a wireless networking card and web interface for controlling and capturing data from the RSI. A combination of two lasers was used as a source for experimental verification. Finally I describe the design and construction of an infrared RSI.

SPATIO-SPECTRAL TRIANGULATION OF VISIBLE AND INFRARED POINT SOURCES USING A PORTABLE ROTATIONAL SHEAR INTERFEROMETER

Jason Richard Gallicchio, MS
Department of Electrical Engineering
University of Illinois at Urbana-Champaign, 2001
David Brady, Adviser

I show how a rotational shearing interferometer (RSI) can determine the location and spectral radiance of a sparse array of point sources. The approach is particularly applicable to situations in which one seeks to estimate a relatively compact set of target positions and spectral components, rather than a pointwise spatio-spectral data cube. In these situations, RSI triangulation provides efficient aperture limited source localization and spectral analysis. I describe a mobile version of an RSI module with built-in electronics and a computer with a wireless networking card and web interface for controlling and capturing data from the RSI. A combination of two lasers was used as a source for experimental verification. Finally I describe the design and construction of an infrared RSI.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF FIGURES

vii

viii

CHAPTER 1

INTRODUCTION

Conventionally, imaging systems seek to to implement pointwise mappings from a source space onto a measurement space. Since hyperspectral systems may involve five dimensional spatial, spectral and temporal source spaces, the sensor data load in such systems is often enormous and untenable. Here, I'll explore optical design and sensor head processing to reduce this data load for sparse source spaces. Our sources consist of one or a few point sources distributed in five dimensional space.

For such sources, acquisition of the full pointwise source map is enormously inefficient. To avoid this data load, one seeks to build sensors that allow efficient computation of relevant source variables, such as position and spectral radiance, without building a full reconstruction of the source space.

I focus on the rotational shear interferometer (RSI). Originally used to measure abberations [1], the RSI mixes spectral and spatial source data on a single sensor plane. The recent trend toward ubiquitous computational power and better sensors for both visible and infrared light make it possible to take advantage of the way the RSI projects the source space onto a sensor plane. Computational analysis can be used to

1

process the RSI sensor's data in many ways. My colleagues in the Photonic Systems Group have previously shown that RSI imaging can be used for tomographic 3D reconstruction and spatio-spectral imaging [2]. Previously, Itoh, Ihoue, and Ichioka [3, 4, 5, 6] constructed an RSI for the purpose of taking a 64x64 image with 64 channels of spectral information. This was done by dithering one arm of the interferometer and capturing 64 two dimensional 64x64 samples of the mutual coherence measured on the CCD. The spectral image was retrieved by performing a 3D Fourier Transform on the 64x64x64 dataset.

As a simplification to such a large dataset, I have used the RSI to measure a particular projection of this spatio-spectral space particularly suited for a sparse array of point sources. In particular, each point maps to a streak on the sensor plane with the angle of the streak being the direction to the source, and the intensity along the streak being the spectrum. A second measurement can be taken at a different angle to numerically calibrate these measurements. In this thesis, simple algorithms are developed that are more appropriate for estimation of these source parameters in real-time situations with a low-power on-board processor.

Coherence imaging, as implemented on the RSI, has many potential advantages over focal imaging, including robustness against abberation and high depth of field. In conventional systems there is a trade-off between resolution and depth of field (the aperture must be reduced to obtain high depth of field, which reduces resolution — the pinhole camera being an extreme example.) RSI systems obtain unlimited depth of field at all apertures, thereby removing this trade off.

2

The trade off in coherence systems is instead between source complexity and signal to noise. Complex sources produce high interferometric shot noise, degrading the reconstructed image. One may stop down the effective aperture in an RSI to degrade source resolution and decrease the number of effective source channels, thereby establishing a tradeoff between resolution and source complexity.

In view of this trade-off, coherence imaging is most attractive for tracking low complexity sparse sources. In this application, the combination of resolution and depth of field is unparalleled. Tracking and analysis of arrays of flying point sources is an example application for this approach. I will first describe the theory behind the operation of the RSI, with emphasis on determining the location and spectrum of point sources. I will then describe the construction of the Portable RSI, which combines a stable RSI with the electronics, processing, and wireless networking necessary to use it as a stand-alone RSI data acquisition system. In the fourth chapter, I discuss experimental results of using the visible RSI for determining the location and wavelength of a laser spot made by combining two different lasers onto the same point. Then I will motivate the use of infrared (IR) with a chapter on IR theory as it applies to the RSI. Finally I will describe how the Portable RSI was used for tracking IR sources.

CHAPTER 2

RSI POINT SOURCE THEORY

2.1   Introduction

The fringe pattern recorded on the RSI's CCD changes in the same way if the object's angle with respect to the optical axis doubles or if the wavelength is cut in half, which will be proven and experimentally demonstrated in the sections to come. It is possible to determine a wavelength-calibrated power spectral density by rotating the interferometer a known amount and taking a second image.

Here I postulate that the source is constrained to be a point radiating at a specific wavelength and explain a scheme for determining the angular position and wavelength in just two 2D CCD measurements. We then extend this scheme to include point sources with discrete and continuous spectra and then sparse arrays of such sources.

This treatment is by no means as general as those dealing with measurement of the 3D mutual coherence function, but the simplification in both hardware accuracy and amount of computation makes this scheme useful where its assumptions are valid. This is especially true in the infrared where one is often very concerned with the

4

spatial and spectral information in a scarcely populated scene using relatively noisy detectors.

2.2 RSI Theory and Design

For simplicity we first consider a simple point source emitting at a single wavelength λ. The phase of the complex analytic field changes by 2π when measured one wavelength farther away in any direction. The complex analytic field emitted by this point source is a spherical wave [7] and is given in equation (2.1) where r is the distance to the point source.

$$E(r) = \frac{A}{r}e^{i2\pi\frac{r}{\lambda}} \tag{2.1}$$

In the RSI pictured in Fig. 2.1, the source is represented as a teapot for visual clarity, whereas the source in this discussion is a single point. The field at the CCD is the sum of the fields reflected off the mirrors in each arm.

We assume the amplitudes of the two fields are approximately the same at the CCD and from now on I neglect the overall amplitude scale. With r_1 and r_2 being the distances from a pixel of the CCD to the point source through the first and second arms of the interferometer, the field at the CCD through the RSI becomes

$$E(r_1, r_2) \propto e^{i2\pi\frac{r_1}{\lambda}} + e^{i2\pi\frac{r_2}{\lambda}} \tag{2.2}$$

The normalized intensity measured is then given by the magnitude squared of this field — a constant plus a co-sinusoidal modulation term.

$$I \quad \propto \quad \left| e^{i2\pi\frac{r_1}{\lambda}} + e^{i2\pi\frac{r_2}{\lambda}} \right|^2$$

5
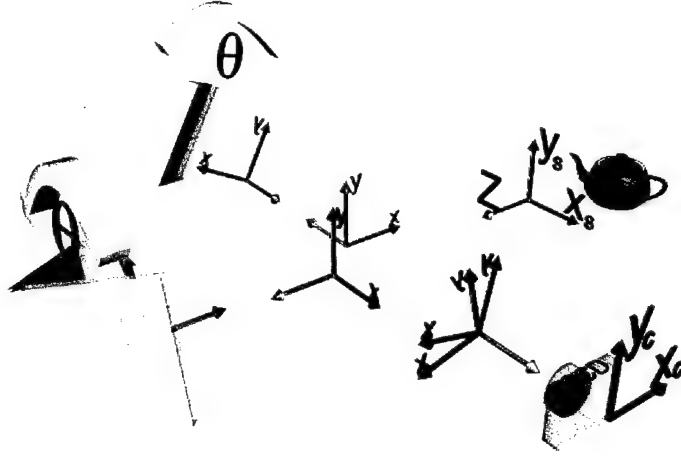
Figure 2.1: RSI with Right Angle Mirrors. Note handedness of reflections.

$$\propto \left(e^{i2\pi\frac{r_1}{\lambda}} + e^{i2\pi\frac{r_2}{\lambda}}\right)\left(e^{-i2\pi\frac{r_1}{\lambda}} + e^{-i2\pi\frac{r_2}{\lambda}}\right)$$

$$\propto 2 + e^{i\frac{2\pi}{\lambda}(r_1 - r_2)} + e^{-i\frac{2\pi}{\lambda}(r_1 - r_2)} \tag{2.3}$$

$$\propto 2 + 2\cos\left(\frac{2\pi}{\lambda}(r_1 - r_2)\right) \tag{2.4}$$

The geometry of the RSI determines r_1 and r_2. Instead of an RSI, to begin with, consider a Michelson Interferometer as shown in Fig. 2.2 with flat mirrors instead of the RSI's right angle mirrors.

Specifically, we want to calculate the intensity measured on a CCD pixel at (x_c, y_c) for a point source at (x_s, y_s) a distance z away where z includes the distance through the arms of the interferometer. These measurements are labelled in Fig. 2.2. By "unfolding" the interferometer, one can see that these distances are given by the simple pythagorean relations in equations (2.5) and (2.6).

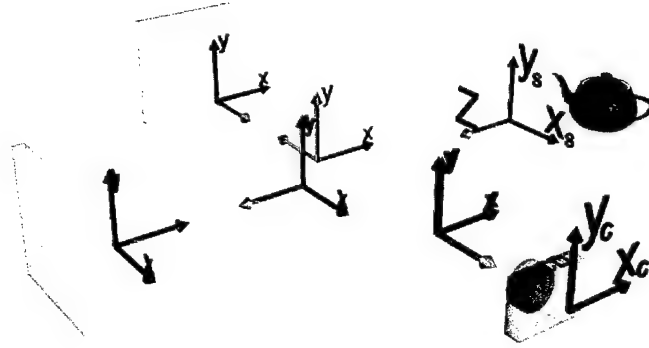$$r_1 = \sqrt{(x_c - x_s)^2 + (y_c - y_s)^2 + z^2} \tag{2.5}$$

6

Figure 2.2: Picture of 3D Michelson Interferometer

$$r_2 = \sqrt{(x_c - x_s)^2 + (y_c - y_s)^2 + z^2} \tag{2.6}$$

Now consider the effect of the right angle mirrors as shown in the RSI pictured in Fig. 2.1, one rotated by angle θ and the other by angle $-\theta$. When you look at yourself in a right angle mirror, you look like you've been flipped across the right angle vertex of the mirror. If the right angle was vertical ($\theta = 0°$), your image will the flipped right-to-left as compared to your image in a flat mirror. Text through the mirror will be readable since the double reflection preserves handedness. If you looked at yourself through right angle mirrors with the right angle horizontal ($\theta = 90°$), you would look upside-down (rotated by $180°$). For an arbitrary angle, you look rotated by 2θ, which is somewhat intuitive since when we turn the prism from $\theta = 0°$ to $\theta = 90°$, we have to go from right-side-up to up-side-down in some continuous way. Fig. 2.3 is a picture

taken through a right angle mirror at an angle of $\theta = 45°$ and thus a "rotation" of $2\theta = 90°$. Note the non-inverted text of "Kodak" on the camera.

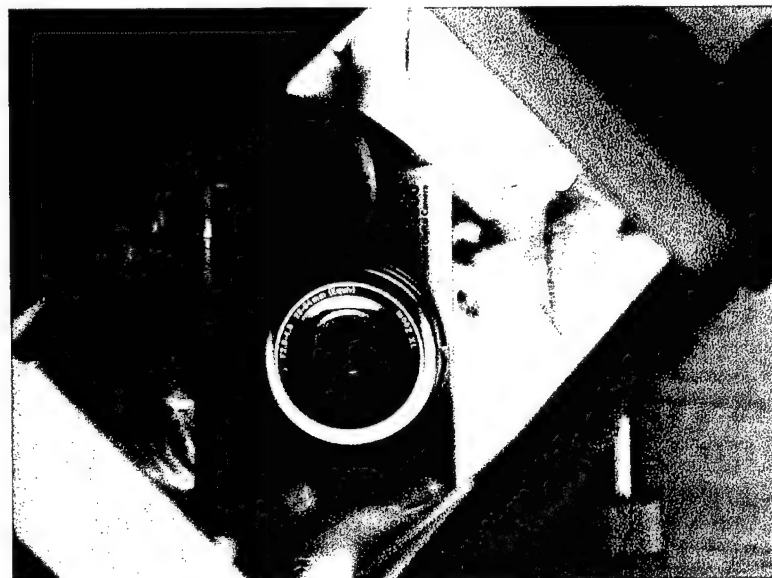Figure 2.3: Looking into a mirror at 45°.

We now consider the affect these right angle mirrors have of fields coming from the source. In Fig. 2.1, a right-handed coordinate system is assigned to the source with the z axis pointing in the direction of light propagation. The effect of the beam splitter and mirrors on the source coordinates is shown after each optical element for both paths. A reflection off of the beam splitter changes the handedness of the coordinate system, whereas a reflection off of a right angle mirror changes the direction of propagation and rotates around the z axis, but through reflections off of each face, it restores the handedness of the coordinate system. The overall effect can be seen in a picture through an actual RSI as shown in Fig. 2.4 where one mirror is vertical and the other mirror is slightly rotated.
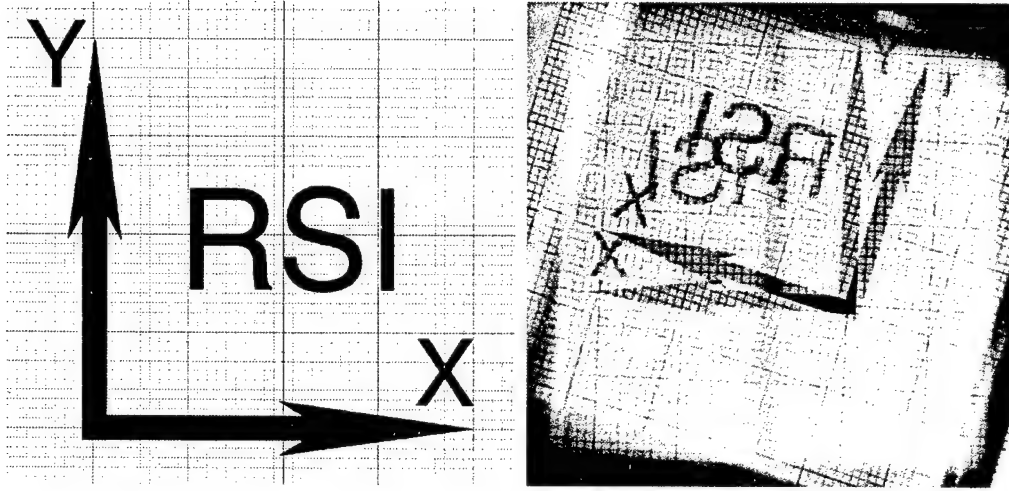
Figure 2.4: Picture of a "Source", and this source through the RSI.

To write this down analytically, we first consider having no shear ($\theta = 0°$) where the right angle of the mirror runs along the y-axis. If you look through either arm of the interferometer, the right angle mirror has the effect of making the source at (x_s, y_s) appear at the same place as a point at $(-x_s, y_s)$ would be through the Michelson Interferometer. We define (x_1, y_1) in the following way: When we look at a point source at (x_s, y_s) through arm 1 of the RSI, it looks like it would be at (x_1, y_1) if we had been looking through a Michelson Interferometer. Similarly, if we were looking through arm 2, the point source looks like it's at (x_2, y_2). For no sheer, $(x_1, y_1) = (x_2, y_2) = (-x_s, y_s)$. For arbitrary shears $-\theta$ in arm 1 and θ in arm 2,

$$
\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} -\cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & \cos(2\theta) \end{pmatrix} \begin{pmatrix} x_s \\ y_s \end{pmatrix}
\tag{2.7}
$$

$$
\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} -\cos(2\theta) & -\sin(2\theta) \\ -\sin(2\theta) & \cos(2\theta) \end{pmatrix} \begin{pmatrix} x_s \\ y_s \end{pmatrix}
\tag{2.8}
$$

9

The quantities we are interested in are r_1 and r_2 given in equations (2.5) and (2.6), but with (x_s, y_s) as seen through the flat mirror interferometer replaced by (x_1, y_1) and (x_2, y_2) due to the effect of the rotated right angle mirrors:

$$r_1 = \sqrt{(x_c - x_1)^2 + (y_c - y_1)^2 + z^2} \tag{2.9}$$

$$r_2 = \sqrt{(x_c - x_1)^2 + (y_c - y_1)^2 + z^2} \tag{2.10}$$

To find the intensity of light on the CCD, we substitute these values for r_1 and r_2 into equation (2.4):

$$I \quad \propto \quad 2 + 2\cos\left(\frac{2\pi}{\lambda}(r_1 - r_2)\right) \tag{2.11}$$

$$\propto \quad 2 + 2\cos\left(\frac{2\pi}{\lambda}\left(\sqrt{(x_c - x_1)^2 + (y_c - y_1)^2 + z^2} - \sqrt{(x_c - x_2)^2 + (y_c - y_2)^2 + z^2}\right)\right) \tag{2.12}$$

At this point, we make the approximation that the distance to the source z is much farther than the distance that the object is off axis:

$$(x_c - x_s)^2 + (y_c - y_s)^2 \ll z^2 \tag{2.13}$$

A Taylor Expansion of the first square root in equation (2.12) for this large z with only the first order terms kept is shown in equation 2.14. A similar expression holds for the second square root.

$$\sqrt{(x_c - x_1)^2 + (y_c - y_1)^2 + z^2} = z + \frac{(x_c - x_1)^2 + (y_c - y_1)^2}{2z} + \cdots \tag{2.14}$$

Substituting this approximation back into equation (2.12):

$$I \propto 2 + 2\cos\left(\frac{2\pi}{\lambda}\left(z + \frac{(x_c - x_1)^2 + (y_c - y_1)^2}{2z} - z - \frac{(x_c - x_2)^2 + (y_c - y_2)^2}{2z}\right)\right) \tag{2.15}$$

10

Now using equations (2.7) and (2.8) to substitute for (x_1, y_1) and (x_2, y_2) and doing trigonometry simplification, we arrive at the RSI point source equation for the intensity measured on a CCD pixel at (x_c, y_c) for a point source at a distance z, radiating with wavelength λ, and for a shear angle set at θ:

$$I \propto 2 + 2\cos\left(\frac{4\pi \sin(2\theta)}{z\lambda}(x_s y_c + y_s x_c)\right) \tag{2.16}$$

Since we are only interested in the direction to the source and not its absolute position, we consider ϕ_x and ϕ_y, the angles from the z-axis of the RSI out to the point source in the x and y directions. With $\tan(\phi_x) = x_s/z$ and $\tan(\phi_y) = y_s/z$, the expression for intensity in equation (2.16) becomes

$$I \propto 2 + 2\cos\left(\frac{4\pi \sin(2\theta)}{\lambda}(\tan(\phi_x)y_c + \tan(\phi_y)x_c)\right) \tag{2.17}$$

2.3   Numerical Simulation of RSI's CCD

As an example, for a point source located 10m away at $(20.5\text{mm}, 30.5\text{mm})$, mirror shear angle $\theta = 10°$, wavelength $\lambda = 500\text{nm}$, Fig. 2.6 shows what is measured on a 200 x 200 pixel, 1cm x 1cm CCD along with its 2D Fourier Transform.

2.4   Separating   the   Spatial   and   Spectral Information

As discussed elsewhere, this RSI pattern is a measurement of the mutual coherence. When the goal was to reconstruct an image via the van Cittert-Zernikie Theorem,
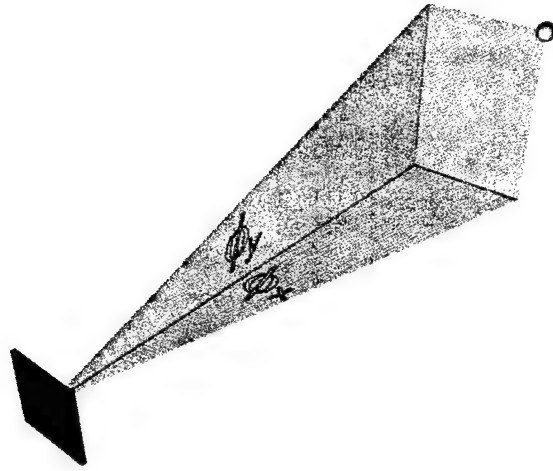
11

Figure 2.5: RSI can only determine direction to source.
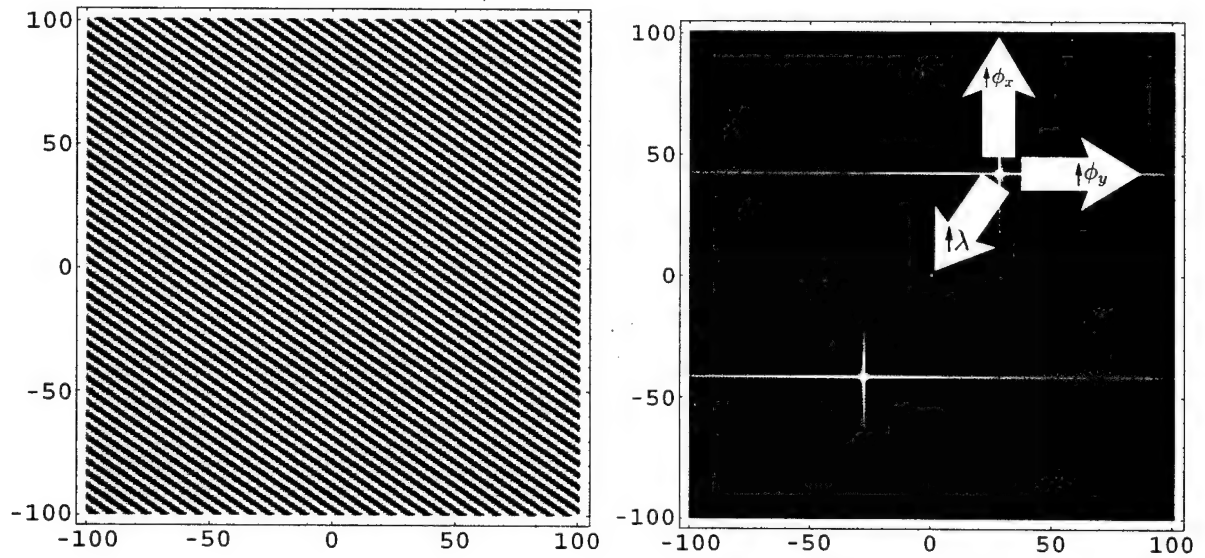

Figure 2.6: Pattern on CCD from point source. Axes labelled in "pixel number" and integer result of FFT.

the Fourier Transform of the CCD intensity gives you projections through the image. This is why the Fourier transform pictured is a single point along with a central DC spot. The DC spot comes from the first, constant term in equation (2.17).

Effects of changing angular position and wavelength are shown in the fourier transform in Fig 2.6. It's apparent from the intensity equation that angular positions are indistinguishable from wavelength in a single measurement. Moving the object twice as far away (thus reducing $\tan(\phi_x)$ and $\tan(\phi_y)$ by a factor of two) has the same effect on the fringe pattern as doubling the wavelength. To break this degeneracy and find the location and wavelength a monochromatic point source, one can first take a measurement with the RSI at some unknown, but desired angle ϕ_x with respect to the source. Then one can rotate the RSI a known angle $\Delta\phi_x$ and take another measurement where the RSI is at angle $\phi_x + \Delta\phi_x$.

How do we know from the fourier transform in Fig. 2.6 if the source was at the top-right or bottom left? Because the data collected from the CCD is purely real, the Fourier Transform is symmetric and there isn't any complex phase data to distinguish the two points in the Fourier Transformed fringe pattern. After the rotation, however, it's obvious which way the point source moved.

The top half of the two dimensional Fourier Transform of the CCD measurements will be a single peak centered around one point in Fourier space measured to be (u_1, v_1) for the unrotated measurement and (u_2, v_2) for the rotated one. From RSI point source equation (2.17), one can see that these "spatial frequencies" correspond to measurements of the angle and wavelength. Note these are not angular frequencies, so the factor of 2π stays inside of the cosine as in $\cos(2\pi v x)$. These spatial frequencies

13

are given by

$$v_1 = \frac{2\sin(2\theta)\tan(\phi_x)}{\lambda} \text{ and } v_2 = \frac{2\sin(2\theta)\tan(\phi_x + \Delta\phi_x)}{\lambda} \tag{2.18}$$

This system of two equations and two unknowns ($\tan(\phi_x)$ and λ) can be solved by expanding $\tan(\phi_x + \Delta\phi_x) = \frac{\tan(\phi_x)+\tan(\Delta\phi_x)}{1-\tan(\phi_x)\tan(\Delta\phi_x)}$, substituting, and taking the negative roots of the resulting quadratic equations since we took the Fourier Transform's positive v_1.

$$\lambda = \frac{\sin(2\theta)}{\tan(\Delta\phi_x)\,v_1 v_2}\left(v_2 - v_1 - \sqrt{(v_1 - v_2)^2 - 4v_1 v_2 \tan^2(\Delta\phi_x)}\right) \tag{2.19}$$

$$\tan(\phi_x) = \frac{1}{2\tan(\Delta\phi_x)\,v_2}\left(v_2 - v_1 - \sqrt{(v_1 - v_2)^2 - 4v_1 v_2 \tan^2(\Delta\phi_x)}\right) \tag{2.20}$$

Though unappealing, this is a closed form for both the wavelength and the direction of the point source using no approximations other than the initial Fresnel approximation from equation (2.13) in deriving the RSI point source equation.

Continuing the numerical example above, if we rotate the RSI by $\Delta\phi_x = 0.1°$, we get the simulated CCD Fourier Transform shown in Fig. 2.7. By a weighted average, the pixel in the Fourier Transform corresponding to the biggest intensity occurs in the first Fourier Transform where v_1 is 28 and in the second where v_2 is 52. Since each pixel in the Discrete Fourier Transform has a spatial bandwidth of one over the size of the CCD,

$$v_1 = \frac{28}{ccdsize} = \frac{28}{10\text{mm}} = 2.8\text{mm}^{-1} \text{ and } v_1 = \frac{52}{ccdsize} = \frac{52}{10\text{mm}} = 5.2\text{mm}^{-1} \tag{2.21}$$

Substituting these values into equation (2.19) and (2.20), one estimates $\lambda = 497.5$nm and $\phi_x = 0.002036 = 20.36\text{mm}/10\text{m}$.
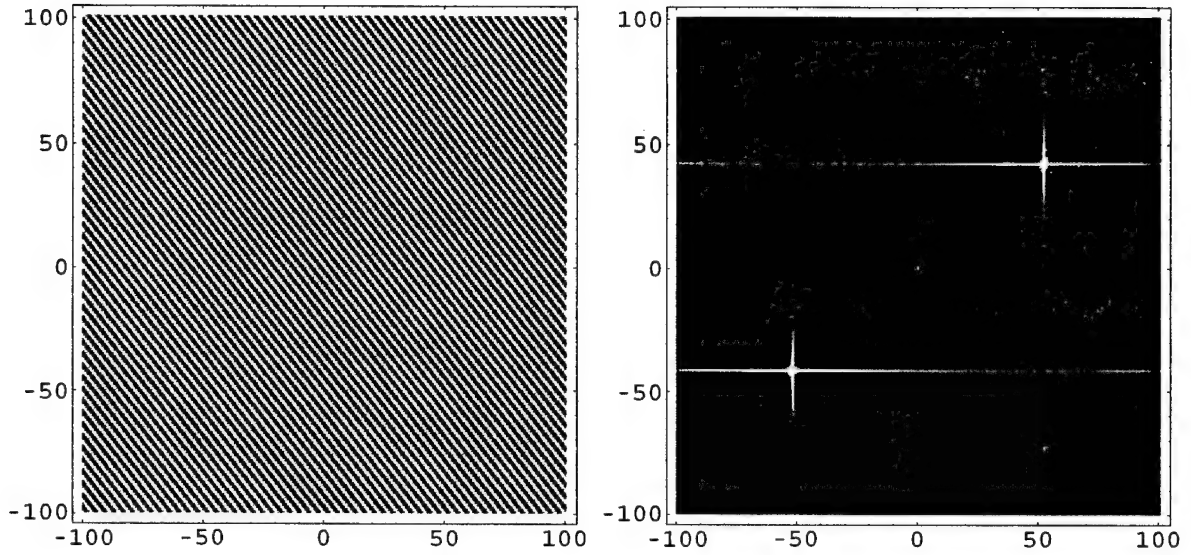
14

Figure 2.7: Pattern on CCD from point source with rotated RSI.

2.5 Spatial and Spectral Range and Resolution

For a single point source, we'll consider "resolution" to be how well one can esti-mate its wavelength (λ) and its angular position (ϕ_x and ϕ_y) and "range" to be the maximum and minimum wavelength and angles that are measurable with a particu-lar shear angle. The resolution and range are determined by the pixel size and pixel count of the CCD.

Let the width and height of each pixel be Δ_x and the number of samples in each direction be N so the total width and height of the CCD is $ccdsize = N\Delta_x$. From the Nyquist Sampling Theorem, the maximum spatial frequency that can be measured without aliasing by the CCD is where maximum and minimum intensities occur in

15

adjacent pixels as in equation (2.22).

$$\cos\left(2\pi u_{max} N\Delta_x\right) = \cos\left(\pi N\right) \tag{2.22}$$

From this way of looking at the maximum frequency that can be sampled, an expression for the maximum measurable spatial frequency is given in equation (2.23).

$$u_{max} = \frac{1}{2\Delta_x} \tag{2.23}$$

Because the spatial frequencies in the FFT image run from $-u_{max}$ to u_{max}, the "size" of each pixel, and thus the accuracy to which it can be measured, is given by Δu in equation (2.24)

$$\Delta_u = \frac{2u_{max}}{N} = \frac{1}{N\Delta_x} = \frac{1}{ccdsize} \tag{2.24}$$

As for range, looking at equation (2.17) for the fringe intensity , u_{max} and Δ_u both apply to the measurement of a combined quantity involving both position and wavelength.

$$\frac{2\sin\left(2\theta\right)}{\lambda}\tan\left(\phi_x\right) < u_{max} = \frac{1}{2\Delta} \tag{2.25}$$

Since both the resolution and range of measurements depend on the shear angle θ, when the shear angle is changed, the effect is to zoom in or out. This can be done to maximize measurement resolution while keeping the point source within range.

2.6 Multiple Sources

If you had a relatively sparse array of point sources at well-defined frequencies, by rotating the RSI in steps and tracking the positions of the point sources, an accurate

16

measurement could be made of the location and wavelength of all of the sources. The key is to group the points (for discrete spectra) or smears (for continuous spectra) on the Fourier transformed RSI data into separate sources. This can be done because as the RSI is rotated around, it will be looking directly at each point source whose data will switch from positive to negative frequency.

CHAPTER 3

DESIGN AND CONSTRUCTION
OF PORTABLE RSI

3.1 Design Criteria

We designed the portable RSI as unit that could be plugged into the wall and automatically run a wireless web-based interface for taking RSI and processing data. This unit is shown in Fig. 3.1 and contains the following:

1. Optical and mechanical elements of a small, stable RSI.

2. CCD camera

3. PC104 Pentium III-based computer with a wireless Ethernet card

4. Control electronics, including

    (a) Power supply

    (b) Rotation stage stepper motor controller

    (c) Piezo translator combined with a position sensor in a feedback circuit to insure linear movement of one arm of the interferometer under computer control.
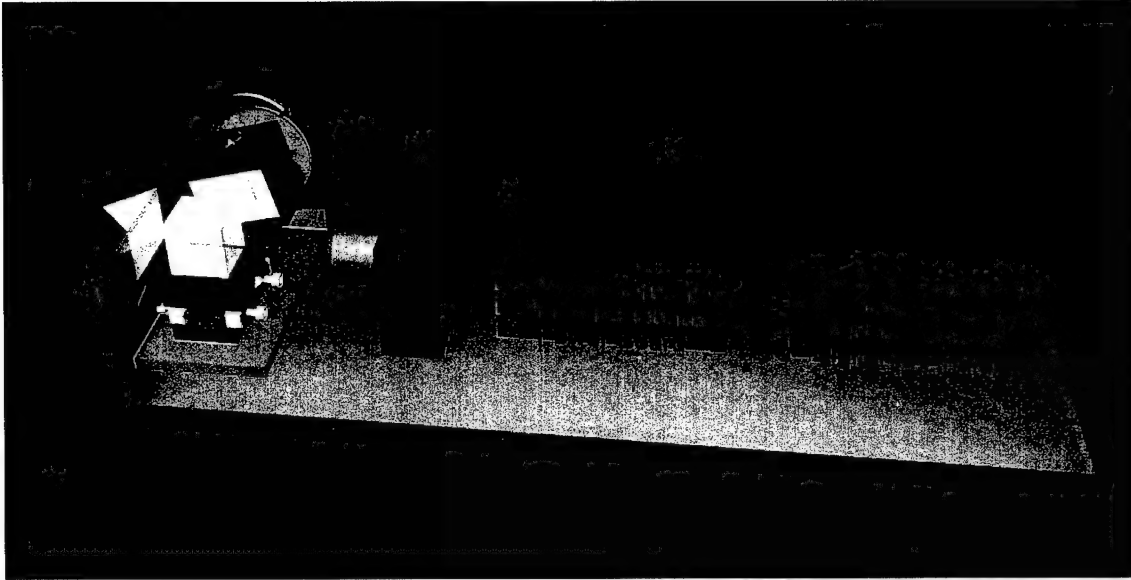
18

Figure 3.1: RSI Module with Computer and Electronics

In one application, a portable handheld iPaq with a wireless ethernet card as shown in Fig. 3.2 runs a web browser that talks to the portable RSI's web scripts. The iPaq is a portable WinCE-based computer with a touch-sensitive screen that using a stylus as input. We use the iPaq to set the shear angle and mirror position, take images, view reconstructions, and organize files. The data continues to be stored on the PC104 computer inside the portable RSI, so when further analysis is required, the data can be downloaded to a desktop PC through the same web interface.

3.2 Optical and Mechanical Elements

The portable RSI is designed to hold optics for both visible light and infrared centered around 10μm. Because the only right angle prisms available for IR are 1in high, this is size we chose for all of the optics. For visible light, we use a cube beam
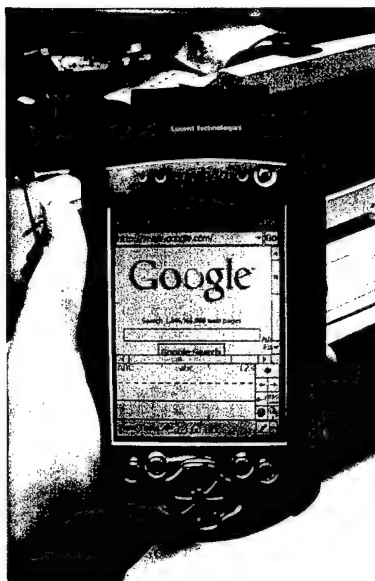
Figure 3.2: An iPaq running a web browser

splitter with an anti-reflection (AR) coating. The high precision right angle prisms are AR coated on the hypotenuse and aluminum coated on the legs. All visible optics were purchased from Melles Griot. There are no cube beam splitters for 10μm IR, so a custom plate beam splitter was constructed and mounted on a quartz square by Janos Technologies, from whom the Aluminum coated IR right angle prisms were also purchased.

Ron Stack and Chuck Henderson contributed significantly to the design of the mechanical elements. In order to keep temperature fluctuations from affecting the alignment of the RSI, the metal alloy Invar was used wherever possible due to its superior thermal stability. All of the parts were machined by the ECE machine shop.

One goal of the portable RSI was to have accurate computer control over the shear angle. This had never been done in any of the interferometers that our group

20

had constructed. Typically rotating one of the right angle prisms to change the shear angle is done by hand and required realigning parts of the RSI afterward. Having automatic rotation without recalibration required extremely precise alignment of the optics. In addition to being highly accurate, the rotation stage we chose had to be very compact. We decided on the Newport PR50PP rotation stage, which has a stepper motor accurate to 0.010 degrees.

In order to change the path length of one arm of the interferometer, a Piezo-Electric disk translator, Physik Instrumente part P-286.40 is mounted on one side of the rotation stage's rotating screw hole as shown in Fig. 3.3. A thin metal disk is mounted on the other side of the hole and a rod connects their centers. The prism holder is mounted onto that rod and the whole assembly of disk translator, thin metal disk, rod, and prism holder rotates together. Care must be taken to align the rod with the rotation stage's axis of rotation, and thanks to the ECE machine shop, no more than .002in of wobble is present at the end of the rod as the rotation stage goes through 360 degrees.

To achieve the sub-micron alignment of the portable RSI required to keep the interference fringes from drifting when the computer rotates the prism or changes the path length with the Piezo Translator, each adjustment must be made using a micrometer translator rather than "by hand." To achieve portability, however, these bulky screw micrometers could not be part of the final Portable RSI design. An external alignment mount is attached to the RSI and the adjustments are made using its micrometers. When fringes are found that do not drift, the components are carefully locked into place using screws. The dove-tail joint shown toward the bottom
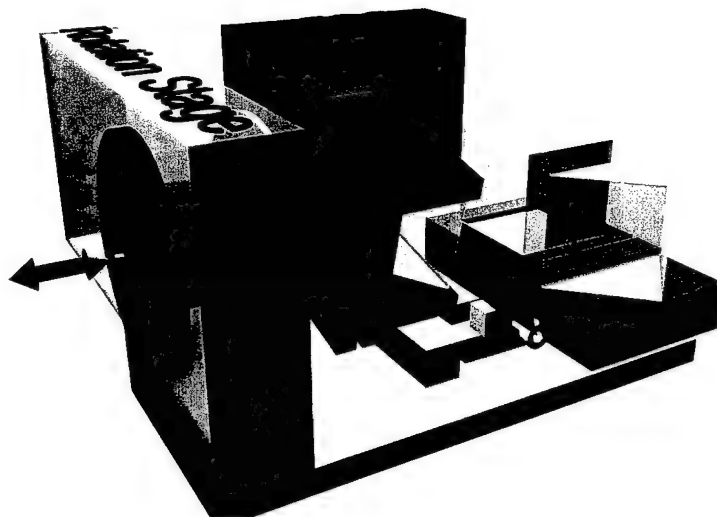
Figure 3.3: Portable RSI, Looking at Rotation Stage and Piezo Disk Translator

of Fig. 3.4 that holds the beam splitter mount to the rest of the RSI leaves room for forward-backward, side-to-side, and twisting motion before being locked into place by the four screws on the sides.

3.3 Alignment

With the beamsplitter and fixed prism assembly (BS/prism holder) removed, a laser is set up level to the table several meters away to shine into the center of the rotating prism. To find the center, translate the laser up and down or back and fourth until it is always centered on the back right angle as the prism rotates around. Next, the tilt of the rotating prism must be set. The goal is to have the beam always retroreflecting into the laser no matter what the rotation is. The complication here is that both the tilt of the prism with respect to the axis of rotation and the axis of
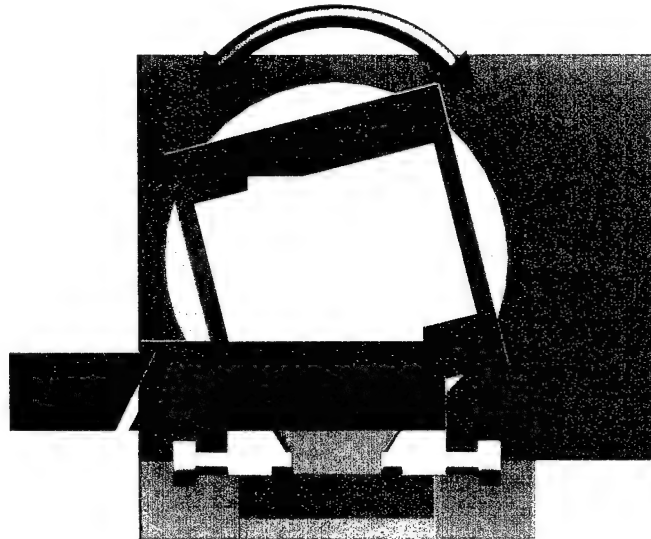
Figure 3.4: Portable RSI, looking at the front. Dovetail joint is on the bottom.

rotation itself must be aligned to the laser. For example, as the prism rotates, if the retroreflection spends more time above the laser, the rotation stage's axis is pointing up. Slips of thin lens paper must be inserted under the back RSI to slightly angle the whole thing down. Similarly, if the spot spends more time to the right of the laser, the whole RSI must be rotated. Once the retroreflected spot makes a roughly even circle around the laser, the adjustment screws on the prism mount itself must be set to reflect the spot back into the laser. The height or position of the laser might have to be adjusted slightly to keep it in the center, and side to side angle along with the slips of lens paper might have to be readjusted to zero in on the correct alignment. When this part is done, the rotation stage should be able to turn 360° while keeping the laser perfectly reflected into itself.

Next, the beam splitter and fixed prism must be aligned to each other. Put a slip of paper in front of the rotating prism to block it from the laser. Unscrew all of the dove-tail screws and push the BS/prism holder against the side of the dove-tail to get it approximately perpendicular to everything else. Put the beam splitter in the mount and rotate it around until the retroreflection off of its faces goes above or below the laser. Then lock it into place with the screws from above. The entire dove-tail slider can be tilted vertically with the screws on the front to get the reflection off of the faces to go into the laser. Slide the BS/prism holder perpendicular to the dove-tail slider so the laser is going through approximately the center of the beam splitter and rotate the assembly (not the BS itself) so that the reflections off of the BS face go into the laser. Lock it into place with the screws.

Next, put the fixed prism in and slide it back and fourth until the laser is centered on the back corner. Rotate the prism back and forth so that the reflection off of its front face goes back into the laser. The reflection off of the anti-reflection coated face will be much dimmer than the reflection off of the Aluminum coated back, so it might be difficult to see.

Now the angles of the RSI are all correct, and all that remains is to slide the BS/prism holder back and forth along the dovetail rail until the path lengths of the two arms are identical to within a few microns. Unfortunately, in order to slide it around, the screws holding it to the dovetail rail have to be loosened and the rotation alignment will be lost. It would be nice if we could find the place where the path lengths are equal using the laser so that at each point the reflections off the face of the beam splitter and fixed prism could be made to point back into the laser. The

best way I know how to find this place is to find the highest contrast fringes for white light. The laser has a very long coherence length, so you'll get high contrast fringes no matter what the path length difference.

Instead of looking at the laser interference, shine white light on an iris with a pin whose point is in the center of the laser beam (right along the optical axis of the RSI.) Turn the laser off and focus on the pin through the RSI with a CCD camera. With the shear angle set to $0°$, the two images of the iris and pin should be on top of each other when everything is aligned properly. Even with the CCD focused on the pin, fringes will still appear when the path lengths are equal. Either by hand or with an external translation stage, move the BS/prism mount back and forth until fringes appear, always keeping the two images of the pin on top of each other. This is the part that can take days. It helps to put an interference filter before the CCD so that fringes appear over a longer distance, but when the highest contrast fringes appear with the filter, they'll also appear with white light. It's a good idea to set the piezo translator to the center of its range to put the zero-path-length-difference there.

3.4   Control Electronics

There were three custom electronics boards used to control the portable RSI. The first is a power supply board that takes an unregulated 24V from either a battery or an AC to DC converter. Using a Lambda PM3024T0512 regulator, this board outputs +5V, +12V, and -12V, which are routed to the PC104 computer and the other electronics boards.
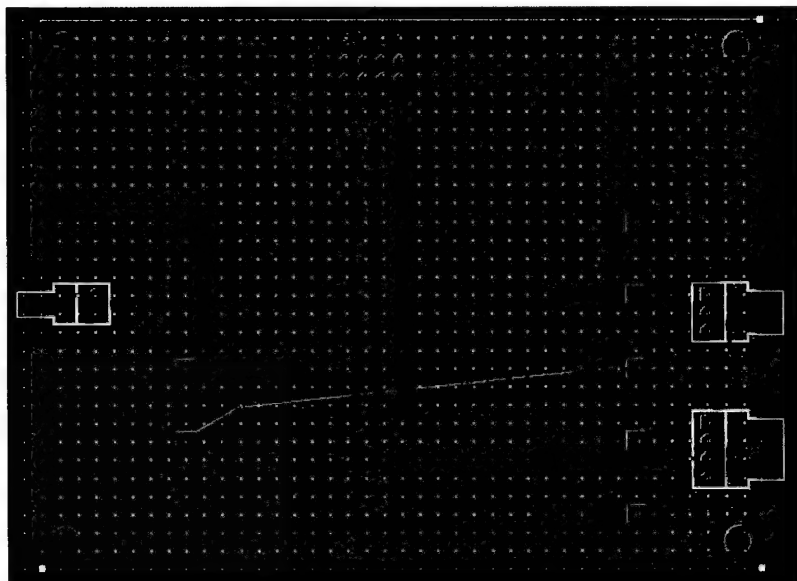
Figure 3.5: Power Supply Printed Circuit Board

The rotation stage stepper motor controller board takes both parallel port data and push-button switches as input. This input is looked at by a Microchip PIC16F84 microcontroller that then generates the logic sequence to activate the two coils of the stepper motor in sequence. These logic levels go to ST PBL3717A stepper motor driver chips that apply 12V to the coils with the appropriate polarity. Using the parallel port, the PC104 computer or an external computer can control the stepper motor to an accuracy of 0.010 degrees.

The purpose of the feedback board is to look at the input and do whatever it has to do to the piezo translator until the sensor matches the desired input. The reason the sensor is used at all is that the position of the piezo disk translaor is not completely linear with respect to the input voltage. It is also subject to historesis,
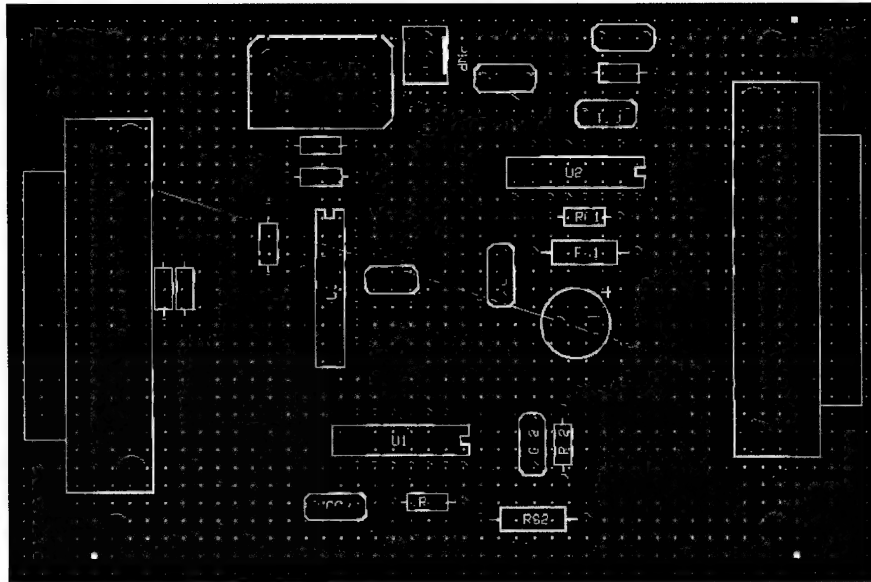
Figure 3.6: Stepper Motor Controller Printed Circuit Board

where after moving around for a while, it won't necessarily return to the same place when the same voltage is applied to it.

A switch mounted on the case selects the input to the piezo feedback board, which takes 0-5V. The first setting selects the PC104 computer's digital to analog (D to A) converter, the second selects an external D to A converter, and the last selects a potentiometer mounted to the case. The feedback board also takes input from a Kaman SMU9000-15N magnetic sensor that outputs 0-12V for positions from 250μm to 350μm away from the aluminum cap on the piezo disk translator. For example, to move to the middle of the range, an input voltage of 2.5V is applied, and the circuit would do whatever it had to do to get the sensor to read 6V. The circuit's output goes to an on-board amplifier unit for the piezo translator that takes an input voltage of 0-5V and amplifies it to the -1000-0V that the piezo requires.
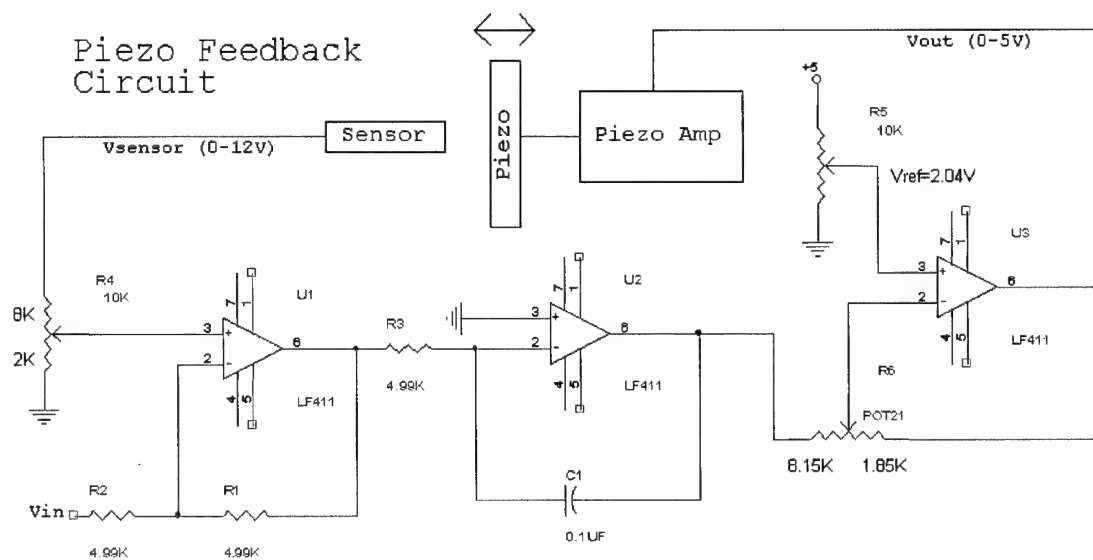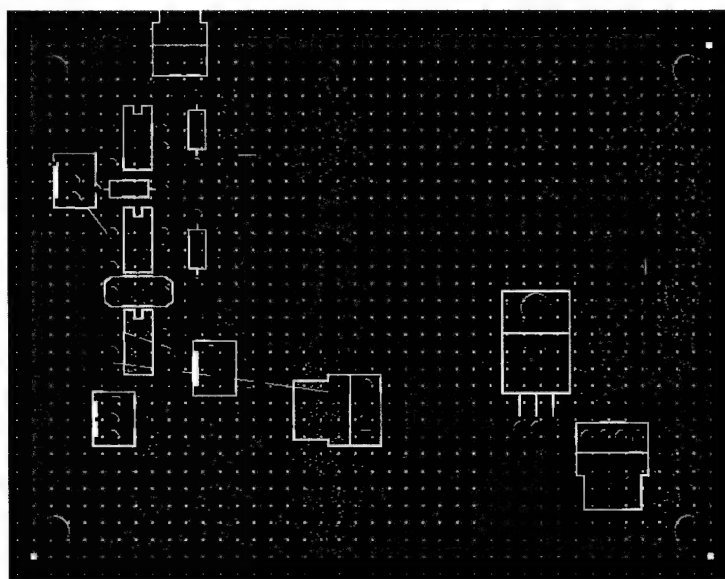
27

Figure 3.7: Piezo Feedback Schematic

Figure 3.8: Piezo Feedback Printed Circuit Board

28

The first part of the circuit calculates the difference between the desired voltage and the current sensor reading by first scaling the sensor's 0-12V to 0-5V and then subtracting it from the input. The second stage is an integrator which integrates all of these voltage differences, gradually increasing or decreasing it's output until the voltage difference is at 0, at which point it holds there. The final stage scales the opamp's approximately -11 to 11V swing to the 0-5V rage required by the on-board integrated piezo voltage amplifier.

3.5   Computer

The computer in the portable RSI unit is a Pentium III PC104 computer form Jumptec. It runs an in-house distribution of Linux which includes the Apache web server for all input-output functionality. It has a wireless Ethernet card and a 256MB compact flash to hold the operating system and the data.

The interfacing is done through CGI scripts accessed through web forms. Those scripts access the parallel port for rotation stage control, the digital to analog converter card for Piezo translation control, and the video capture card to take data. The FFT of the 2D captured fringe pattern and numerical analysis of the data can be performed with results returned through the web interface.

CHAPTER 4

EXPERIMENTAL VERIFICATION

4.1 Setup

We used a beam splitter to combine a HeNe ($\lambda = 632.8$nm) and Argon ($\lambda = 488.1$nm) laser of similar intensity onto rotating diffuser (to eliminate speckle) $1.8m$ away from the RSI at two different ϕ_x angles. A schematic of the setup is shown in Fig. 4.1, and a photo of the experiment is shown in Fig. 4.2 with the rotating diffuser on the left and the RSI on the right.

The camera used was a Photometrics cooled CCD array that was $1in$ by 1in and 1024x1024 pixels, making the pixel size 24.8047μm.

4.2 Results

Fringes captured for two different x positions with $z = 180cm$ in Fig. 4.3.

Matlab's Fast Fourier Transform (FFT) of the fringes is shown in Fig. 4.4. The very lowest frequencies were zeroed out to remove much of the background noise. Even though the lasers were combined into one spot, their different wavelengths caused the
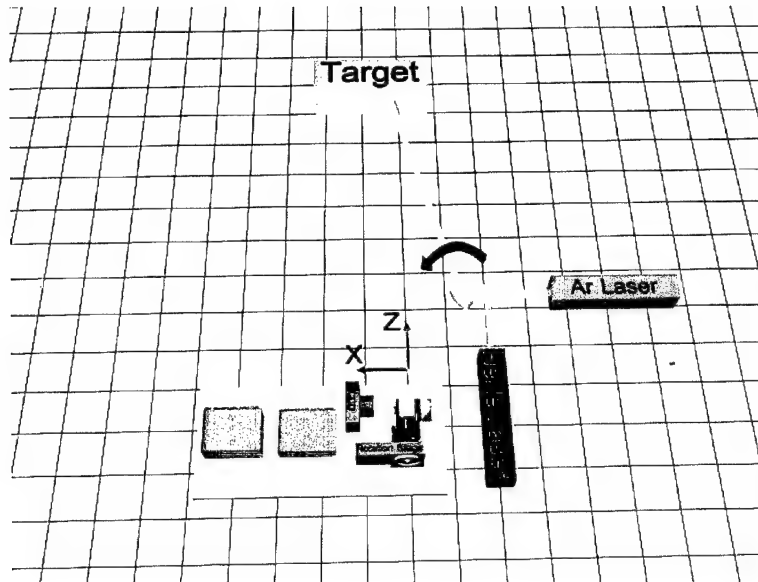
Figure 4.1: Experimental 2-Laser Setup

Figure 4.2: Photo of Experimental Setup

31
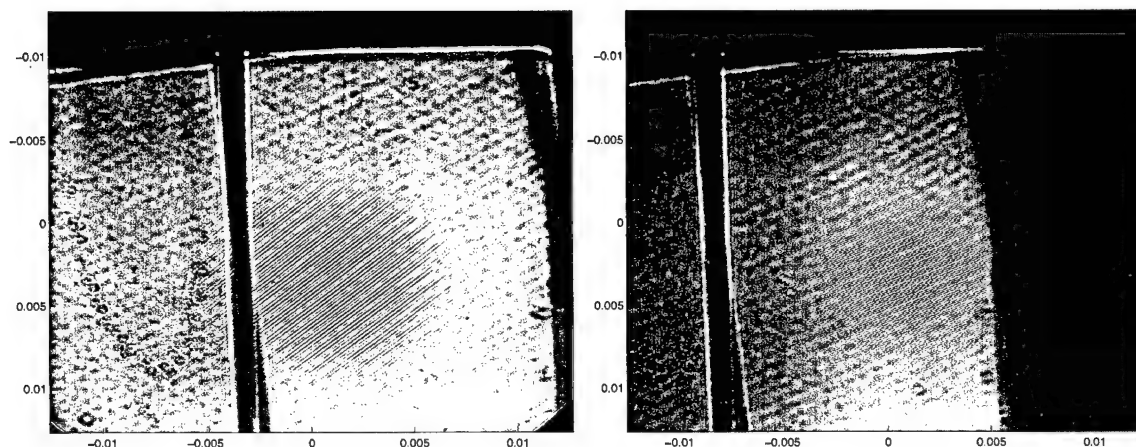
Figure 4.3: CCD capture of point source. (a) $x_s = 4.0$cm (b) $x_s = 9.0$cm

points to appear different distances from the center, but along the same direction. As the source point moved farther to the left, the spots in the FFT moved farther up, which is consistent with equation (2.17) where the x coordinates of the source show up as frequencies in the y direction.
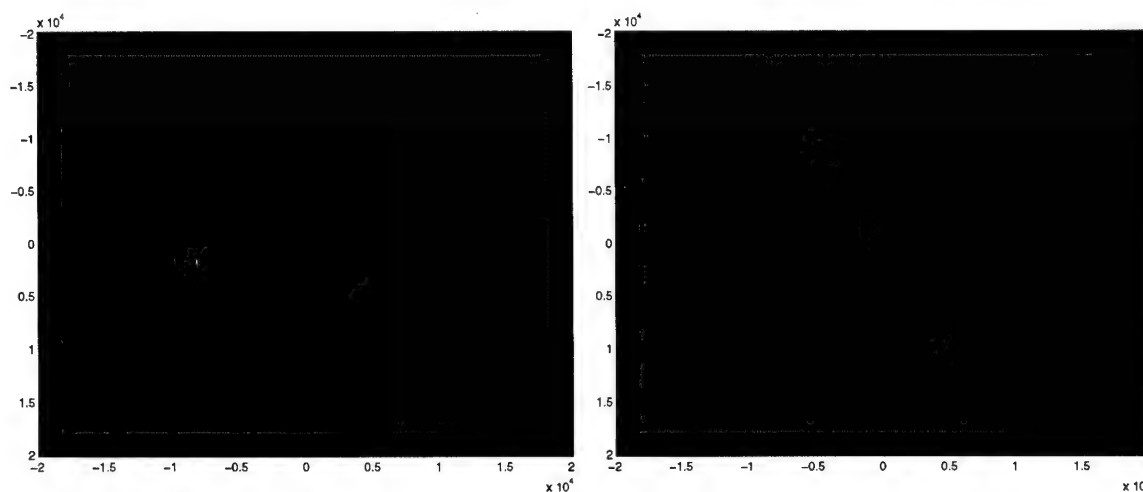

Figure 4.4: FFT of above fringes. (a) $x_s = 4.0cm$ (b) $x_s = 9.0cm$

While all of the theory in this paper was simplified using one mirror rotated θ and the other rotated $-\theta$, for cost and ease of construction, the physical RSI has one fixed mirror and one mirror rotated an angle 2θ. This means that with respect to the hypothetical CCD, the actual CCD is rotated an angle θ, which must be compensated for by rotating either the fringe pattern before fourier transforming, or rotating the data points after they are found in the fourier transform. The shear angle of the RSI was set to $2\theta = 4°$. I chose to rotate the FFT data points since resampling the captured fringe pattern before the fourier transform could introduce errors into the fourier components of the image. When the data is thus rotated, the FFT points in Fig. 4.4 match those predicted by equation (2.17).

CHAPTER 5

INFRARED SOURCES

5.1 Black Body Radiation

Objects at a given temperature emit light with a spectrum roughly of a Black Body with surface properties like reflectiveness causing deviations from this ideal spectrum. The equation [8] for the intensity at a given temperature is:

$$I = \frac{2\pi h c^2}{\lambda^5} \left(\frac{1}{e^{\frac{hc}{\lambda k_B T}} - 1} \right) \tag{5.1}$$

This spectrum is plotted for various temperature in Fig. 5.1. For military and other reasons, it is convenient to look at thermally luminous objects at 10um.

5.2 Radiosity

The study of how a given object with a given temperature distribution emits incoherent light in different directions is called radiosity. The main equation in radiosity [8], equation (5.2), calculates the radiation intensity emitted by a source of area A_s at temperature T a distance r away from a detector with area A_d. The source's normal points an angle θ_s with respect to the vector $\mathbf{r}$ from the source to the detector,
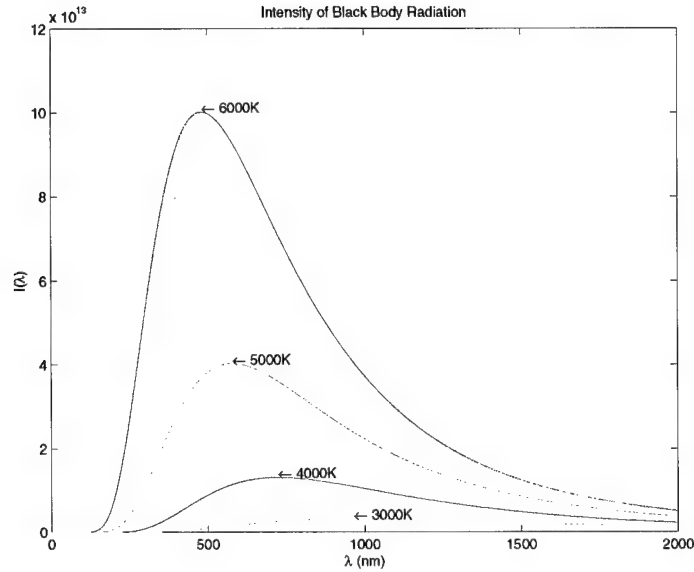
Figure 5.1: Black body spectrum for different temperatures.

and similarly an angle θ_d being the angle from the detector's normal to **r**. These are labelled in Fig. 5.2.

$$I = \frac{A_s \cos(\theta_s) A_d \cos(\theta_d)}{r} \tag{5.2}$$

5.3 Radiosity Calculator

I wrote a program that takes as input a 3D model of an object where each polygon in the model can be assigned a given temperature. The program calculates the intensity of light in a given wavelength range as seen by a simple polygon detector a given distance away and in a given orientation. Fig. 5.3(a) shows the program running with a model of a plane as input, with each surface of the plane at the same temperature.

Figure 5.2: Diagram for Radiosity Equation

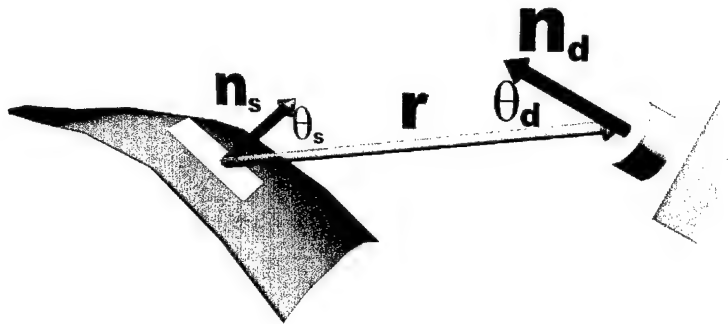Fig. 5.3(b) shows the intensity of light incident on a sphere of triangle detectors in the same orientation as the plane. Notice that more radiation is detected above the flat top of the plane than in front of the nose. This is because more area of the plane faces up than forward.

Figure 5.3: Radiosity Calculation Program: (a) Model (b) Detector

My initial version does not account for obstructions, so in the previous example, if you were looking at the plane in such a way that the wing was blocking part of the body, radiation from both the wing and the body would be counted. It does, however, take into account that polygons facing in the opposite direction should not contribute to the total intensity, so the program is accurate for a convex object where there are no obstructing parts of the model.

In the future, this program will be used to estimate the amount of IR radiation from different objects for use in modelling sources for the IR RSI. A future addition to the program would calculate the fringe pattern seen by the IR RSI's detector for a given source with pieces at given temperatures.

CHAPTER 6

INFRARED RSI

6.1 Construction of the IR RSI

The IR RSI used custom optics from Janos technologies. Their ZnSe right angle prisms were coated on the legs with tin and on the hypotenuse with anti-reflective coating for 10μm. I requested that Janos glue two of these right angle prisms together to form a cube beam splitter, but their engineers claimed there was no glue that was transparent at 10μm. Instead they mounted one of their planar beam splitters on a 1in by 1in square so that it would fit right into the portable RSI. The Portable RSI equipped with the custom prisms and beam splitter is shown in Fig. 6.1

As with the visible RSI, for alignment I illuminated a pin and looked through the IR RSI with and IR camera, moving the beam splitter and prism assembly back and fourth until I saw fringes while always keeping the two images of the pin on top of each other. Fig. 6.2 shows IR fringes with the alignment pin in focus.

For a source, I first tried a 25W power resistor that was painted dark brown. The darker the color, the more radiation it absorbs, and more importantly, the more radiation it emits. This was not bright enough, and from the black body radiation
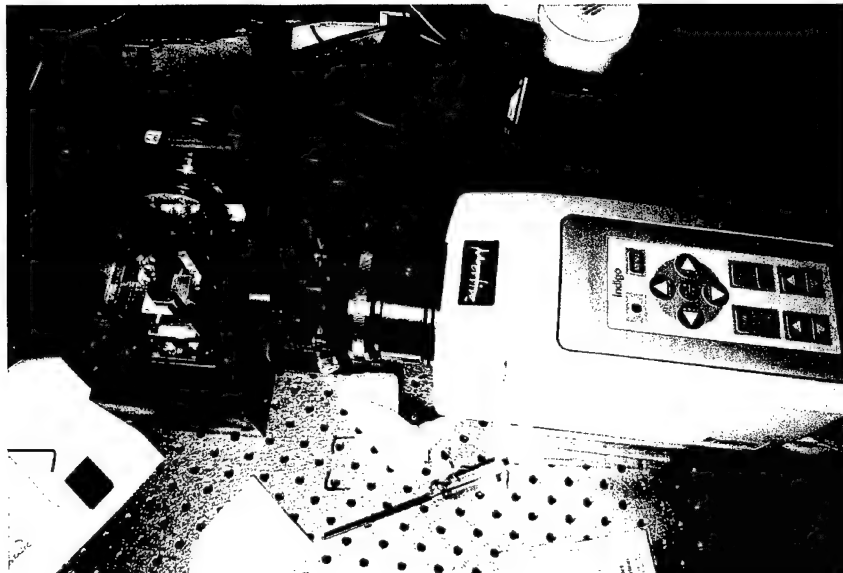
Figure 6.1: Photo of IR RSI
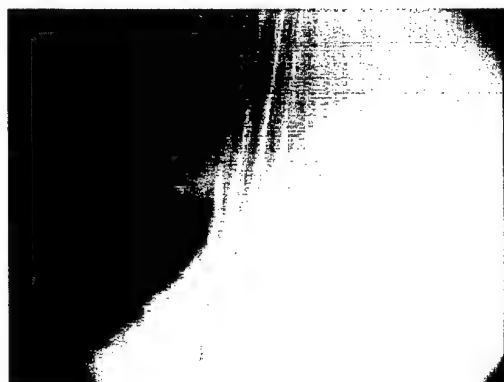
Figure 6.2: IR Fringes with alignment pin

curves in Fig. 5.1, it's clear that as temperature increases, the intensity at all wavelengths continues to increase, even though the peak intensity moves to shorter and shorter wavelengths. For this reason, the filament of a halogen light was used since its temperature is near that of the surface of the sun — around 4500K.

The detector we used a microbolometer array present in Indigo System's Merlin Camera. The camera has a 320 by 200 detector array and both NTSC and 12-bit dynamic range digital output. To capture the digital output, a Bitflow Roadrunner board was used, and custom driver was written for our in-house ImageKitchen video capture scripting software.

CHAPTER 7

CONCLUSION

For the situation where the source is composed of a few points, the RSI is a useful tool for estimating their location and spectra. Since there is no image-forming lens in an RSI system, the location and spectra of both sources that are near and sources that are far can be estimated equally well.

The Portable RSI includes mechanical and optical elements, custom electronics, and a small computer with a wireless networking card. Using a hand-held iPaq computer for control over the data taking process, the Portable RSI simplified experimental verification of my point-source triangulation technique.

With the use of infrared optics and a microbolometer array, information about infrared point sources can be tracked. This is especially useful for tracking distant, thermally luminous sources whose infrared spectra in determine their identity.

REFERENCES

[1] M. Murty, "Interference between wavefronts rotated or reversed with respect to each other and its relation to spatial coherence," *J. Opt. Soc. Am.*, vol. 54, pp. 1187–1190, 1964.

[2] D. L. Marks, R. A. Stack, D. J. Brady, D. Munson, and R. B. Brady, "Visible cone-beam tomography with a lensless interferometric camera," *Science*, vol. 284, pp. 2164–2166, 1999.

[3] K. Itoh and Y. Ohtsuka, "Fourier-transform spectral imaging: retrieval of source information from three-dimensional spatial coherence," *J. Opt. Soc. Am. A*, vol. 3, pp. 94–100, 1986.

[4] K. Itoh, T. Inoue, and Y. Ichioka, "Interferometric spectral imaging and optical three-dimensional Fourier transformation," *Jap. J. Appl. Phys.*, vol. 29, pp. 1561–1564, 1990.

[5] K. Itoh, T. Inoue, T. Yoshida, and Y. Ichioka, "Interferometric multispectral imaging," *Appl. Opt.*, vol. 29, pp. 1625–1630, 1990.

[6] K. Itoh, *Interferometric multispectral imaging*, pp. 145–192. New York: Elsevier Science B.V., 1996.

[7] L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics*. Cambridge: Cambridge University Press, 1995.

[8] E. Hecht, *Optics*. Reading, Massachusetts: Addison-Wesley, 1998.

Three-dimensional coherence imaging in the Fresnel domain

Daniel L. Marks, Ronald A. Stack, and David J. Brady

We show that three-dimensional incoherent primary sources can be reconstructed from finite-aperture Fresnel-zone mutual intensity measurements by means of coordinate and Fourier transformation. The spatial bandpass and impulse response for three-dimensional imaging that result from use of this approach are derived. The transverse and longitudinal resolutions are evaluated as functions of aperture size and source distance. The longitudinal resolution of three-dimensional coherence imaging falls inversely with the square of the source distance in both the Fresnel and Fraunhofer zones. We experimentally measure the three-dimensional point-spread function by using a rotational shear interferometer. © 1999 Optical Society of America

OCIS codes: 030.1640, 110.1650, 110.4850, 100.3010, 100.6890, 070.4550.

1. Introduction

Improvements in electronic sensors, automated positioning systems, and data processing equipment render optical coherence imaging of complex three-dimensional (3D) objects increasingly practical. Two-dimensional (2D) imaging based on the far-field van Cittert–Zernike theorem has been used in radio astronomy for more than two decades.[1] Recently, coherence imaging techniques have begun to shift back to the optical domain,[2,3] and a number of optical systems have been implemented or are under development.[4]

Several researchers have generalized the van Cittert–Zernike theorem to 3D source distributions and have shown that 3D inversion is possible in the far field.[5,6] LaHaie[7] describes modal 3D reconstruction techniques that also work in the near and Fresnel zones. Zarubin notes that the 3D generalized van Cittert–Zernike theorem applies in the Fresnel zone under certain coherence assumptions and that the theorem can also be applied to x-ray and particle scattering.[8] More recently, 3D source re-

construction from a finite far-field aperture by use of the generalized 3D theorem was analyzed and experimentally demonstrated.[9–11] Unlike pseudo-3D techniques such as holography and stereo imaging, coherence imaging provides a true 3D model of object sources.

In this paper we show that Fourier reconstruction techniques can be applied to Fresnel-zone reconstruction by application of a coordinate transformation to the generalized van Cittert–Zernike theorem. This extension is important because the object distance may be much less for a given aperture and wavelength in the Fresnel zone than in the Fraunhofer zone. Because longitudinal resolution falls as the square of object distance, longitudinal resolution in the Fresnel zone may exceed longitudinal resolution in the Fraunhofer zone by several orders of magnitude.

In Section 2 of this paper we review the Fourier-transform relationship between the source intensity distribution and the far-field mutual intensity and describe the coordinate transformation by which a similar relationship is obtained between the source distribution and the Fresnel-zone mutual intensity. In Section 3 we explore the bandpass and resolution limits of 3D coherence imaging. Resolution constraints are easily visualized by use of the 3D spatial bandpass, or band volume, because in limited-aperture systems the band volume has precise boundaries.[12,13] The resolution along any given direction is inversely proportional to the extent of the band volume along that direction. In Section 3 we analyze the band volume and the impulse response for two particular coherence measurement systems,
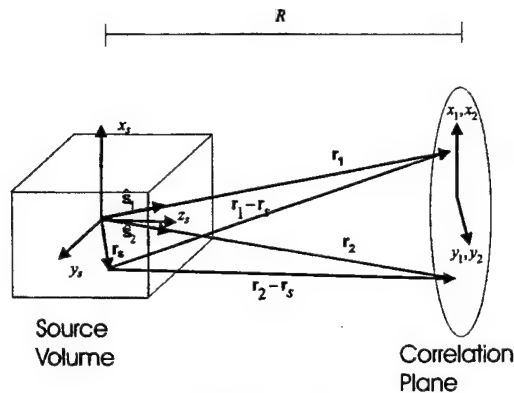
Fig. 1. Measurement geometry for coherence imaging. An incoherent primary source distribution in the source volume is imaged by use of two-point correlation measurements drawn from the correlation plane. The separation between the volumes is greater than the extent of either volume. The correlation point coordinates are (x_1, y_1) for the first point and (x_2, y_2) for the second point. The correlation points are scanned throughout the correlation plane to yield the mutual intensity. $\mathbf{r}_1, \mathbf{r}_2$, vectors from the source volume origin to (x_1, y_1) and (x_2, y_2); $\hat{s}_1, \hat{s}_2$, unit vectors parallel to $\mathbf{r}_1$ and $\mathbf{r}_2$; $\mathbf{r}_s$, position vector in the source volume.

the Michelson stellar interferometer and the rotational shear interferometer. In Section 4 we show experimental reconstructions obtained from an implementation of the rotational shear interferometer.

2. Fourier Inversion of the Generalized Van Cittert–Zernike Theorem

The mutual intensity for a quasi-monochromatic 3D incoherent primary source can be expressed in terms of the source radiant power density by use of the Hopkins integral:

$$
J(\mathbf{r}_1, \mathbf{r}_2) = \left(\frac{k_0}{2\pi}\right)^2 \int_\sigma I(\mathbf{r}_s)
$$

$$
\times \frac{\exp[jk_0(|\mathbf{r}_1 - \mathbf{r}_s| - |\mathbf{r}_2 - \mathbf{r}_s|)]}{|\mathbf{r}_1 - \mathbf{r}_s||\mathbf{r}_2 - \mathbf{r}_s|} \, \mathrm{d}^3\mathbf{r}_s, \quad (1)
$$

where σ is the source volume, $I(\mathbf{r}_s)$ is the 3D source radiant power density, $J(\mathbf{r}_1, \mathbf{r}_2)$ is the mutual intensity at field sample positions $\mathbf{r}_1$ and $\mathbf{r}_2$, and $k_0 = 2\pi/\lambda_0$ is the wave number of the optical field at wavelength λ_0.[14] The geometry of the radiation and measurement space is illustrated in Fig. 1. $\mathbf{r}_s$ is the position vector in the source volume. $J(\mathbf{r}_1, \mathbf{r}_2)$ is measured between pairs of points drawn from an aperture labeled the correlation plane. The correlation plane lies a distance R along the z axis from the center of the correlation volume. As is shown in the figure, $\mathbf{r}_1$ and $\mathbf{r}_2$ are vectors from the origin of the source volume to field sampling points on the correlation plane. $J(\mathbf{r}_1, \mathbf{r}_2)$ is the zero-delay mutual coherence function between the field at $\mathbf{r}_1$ and that at $\mathbf{r}_2$. Systems for measuring $J(\mathbf{r}_1, \mathbf{r}_2)$ are described in section 3.

The goal of coherence imaging is to invert Eq. (1) and reconstruct $I(\mathbf{r}_s)$ from measurements of $J(\mathbf{r}_1, \mathbf{r}_2)$.

Inversion was previously shown to be straightforward in the Fraunhofer zone, where Eq. (1) reduces to the generalized van Cittert–Zernike theorem

$$
J(r_1\hat{s}_1, r_2\hat{s}_2) = \tilde{I}\left[\frac{(\hat{s}_1 - \hat{s}_2)}{\lambda_0}\right] \frac{\exp[jk_0(r_1 - r_2)]}{\lambda_0^2 r_1 r_2}, \quad (2)
$$

where $\hat{s}_1$ and $\hat{s}_2$ are unit vectors in the $\mathbf{r}_1$ and $\mathbf{r}_2$ directions, $r_1 = |\mathbf{r}_1|$, $r_2 = |\mathbf{r}_2|$, and $\tilde{I}(\mathbf{u})$ is the 3D Fourier transform of the source intensity distribution.[5,6] $\mathbf{u}$ is the position vector of $I(\mathbf{r}_s)$ in 3D Fourier space. According to Eq. (2), measurement of $J(\mathbf{r}_1, \mathbf{r}_2)$ over a range in $\mathbf{r}_1 = r_1\hat{s}_1$ and $\mathbf{r}_2 = r_2\hat{s}_2$ yields samples of $\tilde{I}(\mathbf{u})$ for $\mathbf{u}$ over the range $(\hat{s}_1 - \hat{s}_2)/\lambda_0$. Inasmuch as variations in r_1 and r_2 do not affect the range sampled in $\mathbf{u}$, it is sufficient to measure $J(\mathbf{r}_1, \mathbf{r}_2)$ for $\mathbf{r}_1$ and $\mathbf{r}_2$ drawn from a surface surrounding the object rather than a volume. Doing so reduces the six-dimensional measurement space of $J(\mathbf{r}_1, \mathbf{r}_2)$ to four dimensions. Even for $\mathbf{r}_1$ and $\mathbf{r}_2$ drawn from a surface, redundant values of $(\hat{s}_1 - \hat{s}_2)/\lambda_0$ will be obtained. To characterize $I(\mathbf{r}_s)$ to wavelength-limited resolution it is necessary only to sample $J(\mathbf{r}_1, \mathbf{r}_2)$ over a 3D subspace of $\mathbf{r}_1 \otimes \mathbf{r}_2$ that fully samples the range of $\hat{s}_1 - \hat{s}_2$. If $J(\mathbf{r}_1, \mathbf{r}_2)$ is measured over this subspace, $\tilde{I}(\mathbf{u})$ is known for all $\mathbf{u}$ such that $|\mathbf{u}| \leq 2/\lambda_0$. The sphere $|\mathbf{u}| \leq 2/\lambda_0$ is the band volume for this imaging system. The band volume is the 3D spatial bandpass of the imaging system. The impulse response of the system is the inverse Fourier transform of the band volume. For the fully sampled imaging system, the impulse response will be spherically symmetric, with a resolution of approximately $\lambda_0/2$.

In most optical imaging situations, particularly the far-field systems to which Eq. (2) applies, the source volume is remote from the correlation space. In such cases it is not possible to measure $J(\mathbf{r}_1, \mathbf{r}_2)$ over a surface that encloses the source. Rather, the goal for these systems is to reconstruct $I(\mathbf{r}_s)$ from measurements of $J(\mathbf{r}_1, \mathbf{r}_2)$ over a limited range of $\hat{s}_1$ and $\hat{s}_2$. The limited range is illustrated in Fig. 1 by the circular aperture in the correlation plane. The effect of limiting the range of $\hat{s}_1$ and $\hat{s}_2$ is to limit further the range of $\mathbf{u}$ and thereby to reduce the band volume and the imaging resolution.

Rosen and Yariv previously considered source reconstruction from a finite aperture in the far field.[9–11] The result of Rosen and Yariv is obtained by expansion of $\hat{s}_1$ and $\hat{s}_2$ in terms of the planar coordinates of the correlation plane as, for example,

$$
\hat{s}_1 \approx \frac{x_1}{R}\hat{\mathbf{i}}_x + \frac{y_1}{R}\hat{\mathbf{i}}_y + \left(1 - \frac{x_1^2 + y_1^2}{2R^2}\right)\hat{\mathbf{i}}_z, \quad (3)
$$

where R is the distance from the origin of the source volume to the sampling plane. The range over which $\tilde{I}(\mathbf{u})$ is determined from measurements of $J(R\hat{s}_1, R\hat{s}_2)$ in this plane is

$$
\mathbf{u} = \frac{(\hat{s}_1 - \hat{s}_2)}{\lambda_0} \approx \frac{\Delta x}{\lambda_0 R}\hat{\mathbf{i}}_x + \frac{\Delta y}{\lambda_0 R}\hat{\mathbf{i}}_y - \frac{\hat{x}\Delta x + \hat{y}\Delta y}{\lambda_0 R^2}\hat{\mathbf{i}}_z, \quad (4)
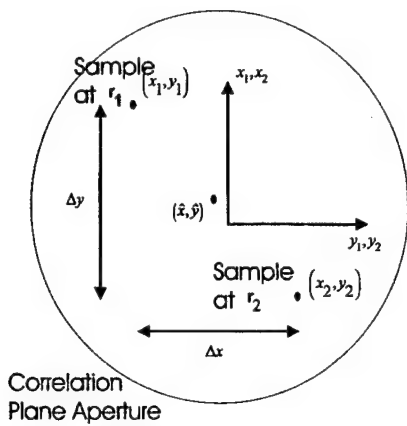$$

Fig. 2. Coordinate geometry in the sampling plane. Source inversion is simplified by use of the transformed coordinates $(\Delta x, \Delta y) = (x_1 - x_2, y_1 - y_2)$, $(\hat{x}, \hat{y}) = [(x_1 + x_2)/2, (y_1 + y_2)/2]$, $q = \Delta x \hat{x} + \Delta y \hat{y}$.

where $\hat{x} = (x_1 + x_2)/2$, $\hat{y} = (y_1 + y_2)/2$, $\Delta x = (x_1 - x_2)$, and $\Delta y = (y_1 - y_2)$. These correlation plane variables are illustrated in Fig. 2. Many different combinations of $\hat{x}$, Δx, $\hat{y}$, and Δy may result in the same $\hat{i}_z$ coordinate in $\mathbf{u}$. To recover all the nonredundant information about $\tilde{I}(\mathbf{u})$ available for a given range in $(\hat{x}, \hat{y}, \Delta x, \Delta y)$ one need only measure the mutual intensity over a 3D projection of the four-dimensional $\hat{x}$, $\hat{y}$, Δx, and Δy space. The 3D projection should sample all allowed values of Δx, Δy, and $\hat{x}\Delta x + \hat{y}\Delta y$. As in Refs. 9–11, we define a variable $q = \hat{x}\Delta x + \hat{y}\Delta y$. We then define the new function $J_{3D}(\Delta x, \Delta y, q)$, which is $J(\mathbf{r}_1, \mathbf{r}_2)$ restricted to the Δx, Δy, q subspace. In this subspace, Eq. (2) becomes

$$J_{3D}(\Delta x, \Delta y, q) = \frac{1}{\lambda_0^2 R^2} \tilde{I}\left(u_x = \frac{\Delta x}{\lambda_0 R}, \right.$$
$$\left. u_y = \frac{\Delta y}{\lambda_0 R}, u_z = -\frac{q}{\lambda_0 R^2} \right). \quad (5)$$

$J_{3D}(\Delta x, \Delta y, q)$ is sampled from two-point correlations over the correlation plane of Fig. 1. The source distribution is recovered from these measurements by inverse Fourier transformation of Eq. (5), which yields

$$I(\mathbf{r}_s) * P_\rho(\mathbf{r}_s) = \lambda R^2 \iiint_\rho J_{3D}(\Delta x, \Delta y, q)$$
$$\times \exp\left(\frac{j2\pi x_s \Delta x}{\lambda_0 R} + \frac{j2\pi y_s \Delta y}{\lambda_0 R} \right.$$
$$\left. -\frac{j2\pi z_s q}{\lambda_0 R^2} \right) d\Delta x d\Delta y dq, \quad (6)$$

where ρ is the range over which the mutual intensity is measured in $(\Delta x, \Delta y, q)$ space and $P_\rho(\mathbf{r}_s)$ is an impulse response for the coherence imaging system. $P_\rho(\mathbf{r}_s)$ is the inverse Fourier transform of the band volume. The band volume in this case is proportional to the sample range in $(\Delta x, \Delta y, q)$. The anal-

ysis leading to Eq. (5) follows discussions in previous publications, especially as presented in Ref. 11.

To derive the Fresnel-zone Fourier relationship we begin again with Eq. (1). Both the Fresnel- and the far-zone approximations rely on paraxial approximations of $|\mathbf{r}_1 - \mathbf{r}_s|$ and $|\mathbf{r}_2 - \mathbf{r}_s|$. The far-field approximation of relation (3), however, includes an assumption that $1/R$ is an accurate approximation of $1/(R - z_s)$ for all points in the source volume at (x_s, y_s, z_s). This assumption severely restricts the transverse extent of the source, as discussed below. Rather than make this assumption, we substitute $1/(R - z_s)$ for $1/R$ in the paraxial approximation. The resultant equations are substantially simpler if we shift the origin of the z axis to the correlation plane. To do this, we define a new variable, $z_{sp} = R - z_s$. The correlation plane then corresponds to the $z_{sp} = 0$ plane, and the paraxial approximation is

$$|\mathbf{r}_1 - \mathbf{r}_s| \approx z_{sp} + \frac{(x_1 - x_s)^2}{2z_{sp}} + \frac{(y_1 - y_s)^2}{2z_{sp}} + \dots . \quad (7)$$

Substituting expression (7) into Eq. (1) yields

$$J(\mathbf{r}_1, \mathbf{r}_2) = \iiint_\sigma \frac{I(\mathbf{r}_{sp})}{\lambda^2 z_{sp}^2} \exp\left[-\frac{j2\pi}{\lambda_0 z_{sp}}(x_s \Delta x + y_s \Delta y) \right.$$
$$\left. + \frac{j2\pi}{\lambda_0 z_{sp}}(\hat{x}\Delta x + \hat{y}\Delta y) \right] d^3\mathbf{r}_{sp}, \quad (8)$$

where $\mathbf{r}_{sp}$ is the position vector in the source coordinates (x_s, y_s, z_{sp}) and, as above, Δx and Δy are the separations between the sampling points and $\hat{x}$ and $\hat{y}$ are the mean positions of the sampling points on the correlation plane. We can obtain Eq. (5) from Eq. (8) if we assume that $1/z_{sp} \approx (1/R)[1 - (z_s/R)]$, such that $(z_s x_s \Delta x)/(\lambda R^2)$, $(z_s y_s \Delta y)/(\lambda R^2) \ll 1$, and that the range of (x_s, y_s) is much less than the range of $(\hat{x}, \hat{y})$. These approximations would mean that the longitudinal extent of the source must be much less than the mean source range and that the transverse extent of the source must be much less than the mean interferometer displacement. These are relatively harsh limitations, particularly in view of the quadratic decrease in range resolution with increasing range.

We can express Eq. (8) as a Fourier transform without making these approximations if we transform the source coordinates into the projective coordinates[15]

$$x' = \frac{x_s}{z_{sp}}, \qquad y' = \frac{y_s}{z_{sp}}, \qquad z' = \frac{1}{z_{sp}} \quad (9)$$

Figure 3 illustrates the transformation between the source coordinates (x_s, y_s, z_{sp}) and the (x', y', z') coordinate system. x' and y' are equal to the tangents of the angles θ_x and θ_y between the ray from the correlation plane's origin to the source point (x_s, y_s, z_{sp}) and to the planes $y_s = 0$ and $x_s = 0$, respectively. In the small-angle approximation, $x' = \theta_x$ and $y' = \theta_y$. Figure 3 is drawn in Cartesian space and shows grids of constant x', y' in the projective space. As indicated by the distortion between the grids, uniform
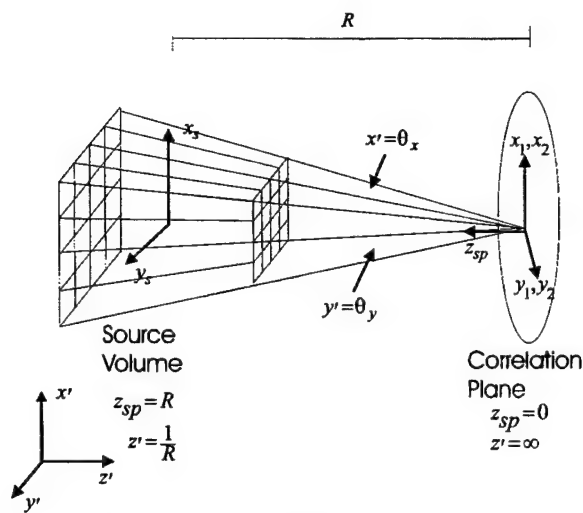
Fig. 3. Relationship between the Cartesian source coordinates and the projective coordinates. The origin of longitudinal coordinate z_{sp} is in the correlation plane. Longitudinal projective coordinate $z' = 1/z_{sp}$ has an origin at $z_{sp} = \infty$ and is equal to $1/R$ at the center of the source volume. In the small-angle approximation, the transverse projective coordinates $x' = z'x_s$ and $y' = z'y_s$ correspond to the angles θ_x and θ_y between the $y = 0$ and $x = 0$ planes and the ray from the correlation plane origin to the real-space source point.

sampling in the primed coordinate system yields non-uniform samples in Cartesian space. The orientation of the primed axes is shown at the lower left of Fig. 3. The origin of the z' axis is at $z_{sp} = \infty$ off the left of the figure. The correlation plane, which is assumed not to lie in the source volume, is at $z' = \infty$.

One can translate displacements in the (x', y', z') coordinate system into real space displacements by using the differential relationships

$$\mathrm{d}x = \frac{z'\mathrm{d}x' - x'\mathrm{d}z'}{z'^2}, \qquad \mathrm{d}y = \frac{z'\mathrm{d}y' - y'\mathrm{d}z'}{z'^2},$$

$$\mathrm{d}z = -\frac{\mathrm{d}z'}{z'^2}. \qquad (10)$$

Neglecting the $\mathrm{d}z'$ dependence of the x and y resolutions and substituting $z = 1/z'$ simplify these relationships to

$$\mathrm{d}x = z\mathrm{d}x', \qquad \mathrm{d}y = z\mathrm{d}y', \qquad \mathrm{d}z = z^2\mathrm{d}z'. \qquad (11)$$

These relationships quantify the distorted Cartesian space grid sampling shown in Fig. 3. For example, uniform steps in $\mathrm{d}z'$ yield grid spaces that increase as z^2 in real space. The sampling resolution decreases in both the transverse and the longitudinal directions as the distance from the image plane increases, but the angular grid spacings $\mathrm{d}x/z$ and $\mathrm{d}y/z$ remain constant.

In the primed coordinates Eq. (8) becomes

$$J_{3D}(\Delta x, \Delta y, q) = \iiint_{\sigma'} \frac{I(x', y', z')}{\lambda_0^2 z'^2}$$

$$\times \exp\left[\frac{-j2\pi}{\lambda_0}(x'\Delta x + y'\Delta y)\right.$$

$$\left. + \frac{j2\pi}{\lambda_0}z'q\right]\mathrm{d}^3\mathbf{r}', \qquad (12)$$

where σ' is the source volume expressed in the transformed coordinates. As in the far-field case, $q = \hat{x}\Delta x + \hat{y}\Delta y$ and $J_{3D}(\Delta x, \Delta y, q)$ is $J(\mathbf{r}_1, \mathbf{r}_2)$ in the 3D subspace. Note that the Jacobian factor for the differential (z'^{-4}) combines with the denominator of source radiation factor (z'^2) to maintain the form of the $1/z'^2$ radiation factor. Equation (12) can be expressed in analogy with Eq. (5) as

$$J_{3D}(\Delta x, \Delta y, q) = \frac{1}{\lambda^2}\tilde{I}_p\left(\frac{\Delta x}{\lambda_0}, \frac{\Delta y}{\lambda_0}, \frac{q}{\lambda_0}\right), \qquad (13)$$

where $\tilde{I}_p(\mathbf{u})$ is the 3D Fourier transform with respect to (x', y', z') of $I(x', y', z')/z'^2$. Since care must be taken to avoid the singularity at $z' = 0$, the Fourier transform of $I(x', y', z')/z'^2$ cannot be taken over unbounded space. The range of integration is bounded by the assumption that the source distribution has finite support and that the correlation plane is far removed from the source. As in the far-field case, one recovers the incoherent intensity distribution of the distributed source by inverse Fourier transforming Eq. (13), using the mutual intensity sampled in the correlation plane in pairs of coordinates parameterized in $(\Delta x, \Delta y, q)$. In analogy with Eq. (6), this approach yields

$$\left[\frac{I(x', y', z')}{z'^2}\right] * P_\rho(\mathbf{r}') = \lambda\iiint_\rho J_{3D}(\Delta x, \Delta y, q)$$

$$\times \exp\left[\frac{j2\pi}{\lambda_0}(x'\Delta x + y'\Delta y)\right.$$

$$\left. - \frac{j2\pi}{\lambda_0}z'q\right]\mathrm{d}\Delta x\mathrm{d}\Delta y\mathrm{d}q. \qquad (14)$$

The source intensity in the Cartesian-space coordinate system is determined by transformation of $I(\mathbf{r}')$ to the $\mathbf{r}_s$ coordinate system. As in the far-zone case, $P_\rho(\mathbf{r}')$ is the Fourier transform of the band volume. The primary difference between Eqs. (5) and (6) and Eqs. (13) and (14) is that the ratio of the source extent to the source range can be larger in the latter. This means that Eqs. (13) and (14) can accurately reconstruct a source of a given size from a closer range than can Eqs. (5) and (6). Because the band volume in both cases scales with the ratio of lateral aperture to the source range, a closer range means a bigger band volume and better resolution.

3. Measurement Systems, Band Volume, and Impulse Response

In Section 2 we derived source-reconstruction algorithms from limited-aperture coherence measurements. In this section we consider two particular physical systems for obtaining these coherence measurements and we analyze the band volume and the impulse response for each system. The two systems that we consider are the Michelson stellar interferometer (MSI) and the rotational shear interferometer (RSI). The MSI, first implemented in 1878,[16] is the prototypical astrometry instrument and was used by Rosen and Yariv to demonstrate 3D coherence imaging. MSI has general utility as an interferometer, but because it collects only one correlation for each instrument position it is extraordinarily inefficient as an imaging instrument. The RSI has a much briefer history but still has been under investigation for more than three decades.[17] Roddier[2] and Roddier and Rodier[18] have used the RSI for 2D imaging, and Itoh et al.[19-22] used RSI data to reconstruct 3D data sets consisting of two spatial dimensions and one spectral dimension. The advantage of the RSI is that it samples entire planes of independent coherence measurements in parallel. In this section we briefly review data acquisition with the MSI and RSI, and we analyze the band volume and the impulse response for typical implementations of each instrument.

The MSI consists of two field-sampling ports mounted upon a single mechanical beam. One combines the field drawn from the sampling ports through an optical system to determine the mutual intensity between the two sampling points. Various mechanisms can be employed to adjust the relative path lengths from the sampling ports to the detector that determines the mutual intensity. The phase and the amplitude of the mutual intensity can be extracted from a spatial fringe pattern or by dithering of the relative delay of the optical paths. We do not consider the beam-combining optics here. It is useful, however, to consider the sample point geometry in the correlation plane, which is illustrated in Fig. 4. The sample points lie upon the mechanical beam that passes through the origin of the correlation plane. The beam rotates freely about the origin but cannot be displaced. The sample points may lie anywhere along the beam. We define the new variable r_{max} to be the maximum distance of a sampling point from the origin. r_{max} is also the radius of the correlation plane aperture and half of the length of the MSI beam.

We define three new variables to describe the state of the MSI. ϕ is the angle between the mechanical beam and the x axis. The range of ϕ is $[0, 2\pi]$. d is the separation between the sampling points. The range of d is $[0, 2r_{max}]$. $\hat{d}$ is the distance of the midpoint between the sampling ports from the origin. The range of $\hat{d}$ is $[(d/2) - r_{max}, r_{max} - (d/2)]$. The
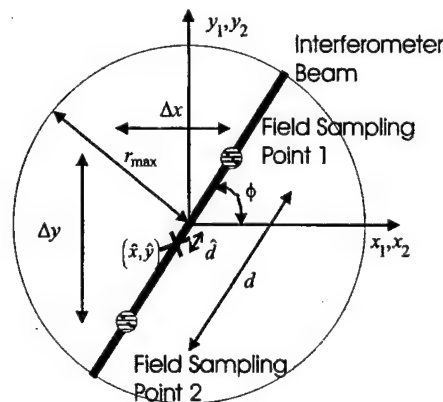

Fig. 4. Geometry of the MSI correlation plane: r_{max}, radius of the system aperture; d, sampling-point separation; $\hat{d}$, distance between the midpoint of the sampling points and the origin; ϕ, angle between the interferometer beam and the x axis.

sample space coordinates for a given interferometer state are

$$\Delta x = d\cos\phi,$$
$$\Delta y = d\sin\phi,$$
$$q = \hat{d}d. \tag{15}$$

Using the ranges described above, we find that Δx and Δy cover the range $[-2r_{max}, 2r_{max}]$. The range of q varies as a function of Δx and Δy. For a given value of $d = \sqrt{\Delta x^2 + \Delta y^2}$, the range of q is $\{-d[r_{max} - (d/2)], d[r_{max} - (d/2)]\}$. The band volume is the range of $\mathbf{u}$ over which $\check{I}(\mathbf{u})$ or $\check{I}_p(\mathbf{u})$ can be sampled. The band volume is the Fourier transform of the impulse response $P_p(\mathbf{r}_s)$, which is used in Eqs. (6) and (14). In the far-field reconstruction of Eq. (5), we find that $\mathbf{u} = [(\Delta x/\lambda_0 R), (\Delta y/\lambda_0 R), (q/\lambda_0 R^2)]$. For the projective coordinates used in the Fresnel case, we find from Eq. (13) that $\mathbf{u} = [(\Delta x/\lambda_0), (\Delta y/\lambda_0), (q/\lambda_0)]$. In both cases the band volume is proportional to the range of $(\Delta x, \Delta y, q)$.

The band volume for the MSI is sketched in Fig. 5. The axes in the figure are scaled in terms of r_{max}/R, where R is the nominal distance from the source to the correlation plane. Because $r_{max}/R < 1$ in both the Fraunhofer and the Fresnel domains and because u_z scales as r_{max}^2/R^2, the u_z axis is greatly expanded relative to the u_x and u_y axes in the figure. The band volume is useful in estimating the resolution of source reconstruction and in designing the coherence sampling scheme. The resolution along any given direction can be approximated by the inverse of the extent of the band volume along that direction. This approximation yields transverse resolution $\lambda_0 R/r_{max}$ and longitudinal resolution $\lambda_0 R^2/r^2_{max}$ in the Fraunhofer zone. In the primed Fresnel-zone coordinates, the transverse resolution is λ_0/r_{max} and the longitudinal resolution is λ_0/r_{max}^2. Note that the transverse and the longitudinal coordinates are not in the same units. When the conversion factors from distorted to Cartesian space listed in Eq. (9) are applied, the Fresnel-zone resolution is also $\lambda_0 R/r_{max}$ in trans-
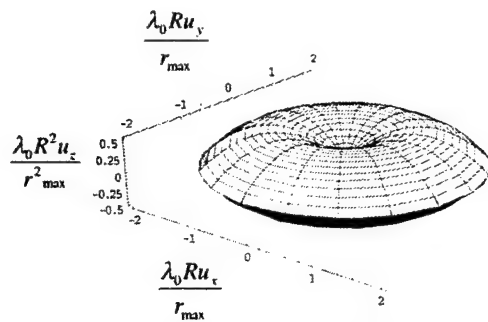
Fig. 5. Band volume for MSI sampling. The band volume is plotted in the real-space Fourier space of the source density for the Fraunhofer zone and in the projective-space Fourier space for the Fresnel zone. The coordinate axes correspond to the Fourier coordinates (u_x, u_y, u_z). The transverse coordinates are normalized with respect to $r_{\max}/\lambda_0 R$. The longitudinal coordinate is normalized with respect to $r_{\max}^2/\lambda_0 R^2$. Because $r_{\max}/R \ll 1$, the normalization frequency for the longitudinal axis is less than it is for the transverse axes. The missing cone in the Fourier space along the u_z axis is characteristic of limited-angle tomographic systems.
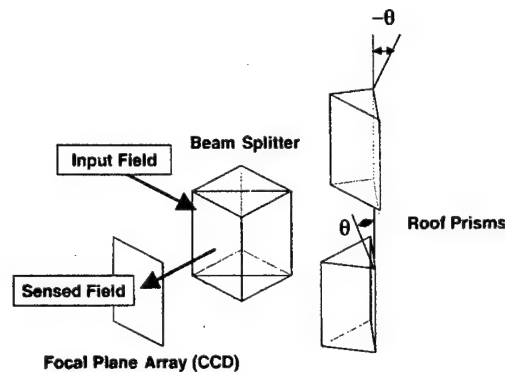

Fig. 6. Basic structure of a rotational shear interferometer. The RSI is a Michelson interferometer in which the plane retroreflection mirrors have been replaced with folding mirrors. The folding axes of the mirrors lie in the transverse plane at angles ϕ and $-\phi$ with respect to the x axis. The output port interferes differentially rotated wave fronts from the two mirrors.

verse coordinates and $\lambda_0 R^2/r_{\max}^2$ longitudinally. These estimates are confirmed in models of the impulse response presented below.

The contraction of the band volume along the longitudinal axis near the origin (the "missing cone"[15]) acts as a high-pass filter on the reconstructed source distribution. This filtering has two effects: First, objects with high-frequency transverse spatial features will be better resolved longitudinally than more nearly uniform objects and, second, interference sampled between distant points covers more of the band volume and thus contains more information than interference between near neighbors. The first effect means that one could not resolve a longitudinally distributed set of uniform planar sources at all but that one could easily resolve discrete point sources. The Fourier-space representation of the planar sources is a set of points on the u_z axis, exactly orthogonal to the band volume. The Fourier representation of the point sources covers the u_x–u_y plane, fully overlapping the band volume. In view of the second point one may seek to measure correlation samples with separations that match the direction of the lobes of the band volume so that high-frequency details that provide depth information are not missed.

As was mentioned above, the MSI samples only one correlation per instrument position. Other designs based on RSI's are attractive because they sample complete planes in $(\Delta x, \Delta y)$ space in parallel. The primary components of the RSI, as illustrated in Fig. 6, are a beam splitter and two folding mirrors. The field incident upon the folding mirrors is reflected back through the output port of the beam splitter and detected at every point in the output plane. The folding mirrors consist of two planar reflectors joined at right angles. The folding mirrors are often implemented by the use of roof prisms. Each folding mirror inverts the reflected field about its fold axis. The

fold axes of both mirrors lie in the transverse plane. As shown in Fig. 6, the fold axis of one mirror makes an angle θ with respect to the x axis. The fold axis of the other mirror makes an angle $-\theta$ with respect to the x axis. Let the transverse coordinates in the output planes of the RSI be (x_f, y_f). The field produced at the output point (x_f, y_f) by the fold mirror with an axis at angle θ relative to the x axis is the field that would appear at $(x_f \cos 2\theta - y_f \sin 2\theta, -x_f \sin 2\theta - y_f \cos 2\theta)$ if the fold mirror were replaced with a plane mirror. If, for example, $\theta = 0$, the fold mirror would reflect across the x axis and the output point (x_f, y_f) would correspond to the plane-mirror output point $(x_f, -y_f)$. The field produced at (x_f, y_f) by the $-\theta$ mirror would appear at $(x_f \cos 2\theta + y_f \sin 2\theta, x_f \sin 2\theta - y_f \cos 2\theta)$ if that mirror were replaced with a plane mirror. The mutual coherence between fields that is due to the two mirrors can be determined by longitudinal dithering of one of the fold mirrors. Each point in the output window samples the mutual coherence for a distinct transverse separation relative to the plane-mirror Michelson interferometer. The separations and mean positions for the mutual coherence sampled at (x_f, y_f) are

$$\Delta x(x_f, y_f) = 2y_f \sin(2\theta),$$
$$\Delta y(x_f, y_f) = 2x_f \sin(2\theta), \tag{16}$$

$$\hat{x}(x_f, y_f, x_g) = 2x_f \cos(2\theta) + x_g,$$
$$\hat{y}(x_f, y_f, y_g) = -2y_f \cos(2\theta) + y_g, \tag{17}$$

where (x_g, y_g) is the transverse displacement between the origins of the output plane coordinates and the source volume coordinates. The q coordinate at each point in the output plane is

$$q = \Delta x(x_f, y_f)\hat{x}(x_f, y_f, x_g) + \Delta y(x_f, y_f)\hat{y}(x_f, y_f, x_g)$$
$$= (y_f x_g - x_f y_g)\sin(2\theta). \tag{18}$$

A RSI samples a surface in $(\Delta x, \Delta y, q)$ space for each value for (x_g, y_g). To sample the entire accessible 3D

Fig. 9. Surface plot of the 3D MSI impulse response in the x'–z' plane. The vertical axis is normalized to the maximum response. The spatial axes are in projective coordinates, with units of inverse meters for the longitudinal axis and radians for the transverse axis. The impulse response is approximately shift invariant in the projective space; it is not shift invariant in real space. To obtain the real-space impulse response one adds $1/R$ to the longitudinal range and takes the inverse. For an impulse at 1 m, a point at $z' = 0.1$ is at $z = 1/(1 + 0.1) = 0.91$. A point at $z' = -0.1$ is at $z = 1/(1 - 0.1) = 1.11$.
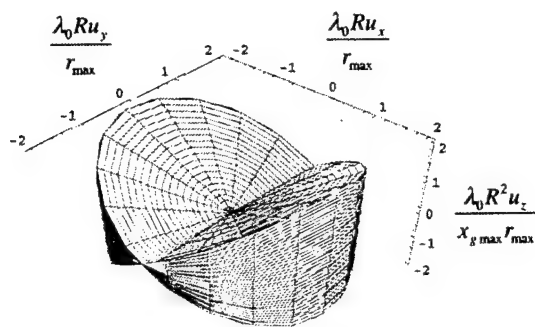
Fig. 7. Band volume for linear translation RSI sampling. The situation is identical to that of Fig. 4, except that the normalization of the u_z axis is now $x_{g_{max}} r_{max}/\lambda_0 R^2$. $x_{g_{max}}$ is the linear displacement range for the RSI. For the RSI of this figure, $\phi = \pi/4$ and the fold axes of the two mirrors are perpendicular. In this geometry, the RSI is also called a wave-front folding interferometer.

$(\Delta x, \Delta y, q)$ space one translates the interferometer in (x_g, y_g).

As for the MSI, let r_{max} be the radius of the RSI aperture. Let $x_{g_{max}}$ be the distance over which the RSI is translated transversely to the optical axis. The band volume captured by a RSI with $\theta = \pi/4$ translated linearly along the x_g axis is shown in Fig. 7. $\theta = \pi/4$ corresponds to a special class of RSI, the wave-front folding interferometer.[23] The crease in the band volume of Fig. 7 along the u_y axis is due to the fact that q vanishes normally to the translation direction for linear translation of the RSI so that only object features and edges perpendicular to the path contribute to longitudinal resolution. This crease can be avoided by translation of the RSI along a nonlinear path. For example, Fig. 8 shows the band volume when RSI with circular aperture r_{max} is translated in a circle of radius $x_{g_{max}}$. Reducing θ decreases the effective aperture of the RSI and thereby decreases its resolution. This may be desirable, particularly if one wishes to match interference fringes in the output plane to a CCD pixel spacing.

We now consider impulse responses for MSI and RSI systems. As a benchmark of the resolution of these systems, we calculate impulse response under
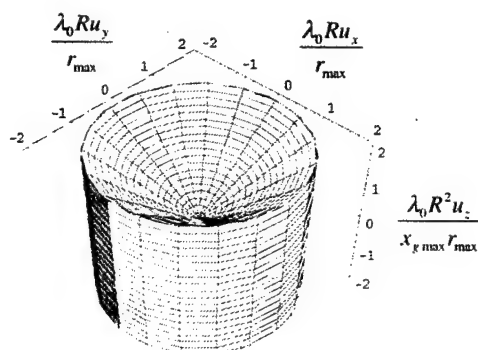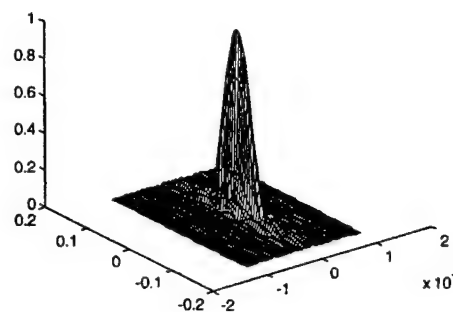

Fig. 8. Band volume for imaging with a circularly translated RSI. The situation is identical to that of Fig. 6, except that $x_{g_{max}}$ now represents the radius of the circle about which the optical axis of the RSI is translated.

the assumption that the correlation plane aperture is fixed in space. This assumption is not likely to reflect practical RSI uses in which the aperture moves with the instrument, but a fixed aperture gives us a common basis for comparing the two interferometers. To find the impulse response for the fixed aperture we set correlations between pairs of points where either pair of correlated points (x_1, y_1) or (x_2, y_2) was outside the aperture $(x_1^2 + y_1^2) > r_{max}^2)$ or $(x_2^2 + y_2^2) > r_{max}^2)$ to zero.

We model the impulse response of a Fresnel-zone imaging system by calculating the mutual intensity as a function of space, using Eq. (13), and then inverting the mutual intensity to find the filtered source intensity, using Eq. (14), under aperture and sampling constraints of the imaging system. Our simulations use a discrete $64 \times 64 \times 64$ point-source volume, consisting of one nonzero intensity point, that was propagated by means of a fast Fourier transform to provide the mutual intensity correlations as a function of Δx, Δy, and q. We then inverse by fast Fourier transform the mutual intensity to reconstruct the filtered source intensity. Because the input is a single point, this reconstruction is the impulse response. In general, each $(\Delta x, \Delta y, q)$ corresponds to multiple pairs of correlation plane points (x_1, y_1) or (x_2, y_2). The MSI and RSI approaches, as well as other potential sampling schemes, improve sampling efficiency by associating each $(\Delta x, \Delta y, q)$ with unique values of (x_1, y_1) and (x_2, y_2).

Figure 9 is a surface plot of a cross section of the impulse response for MSI sampling for a 0.5-cm aperture and a wavelength of 632.8 nm. The lateral coordinate is in angular units and the longitudinal coordinate is in units of inverse distance, consistent with the primed coordinate system. To transform these units into real space, one multiplies the transverse coordinate by the source–correlation plane dis-
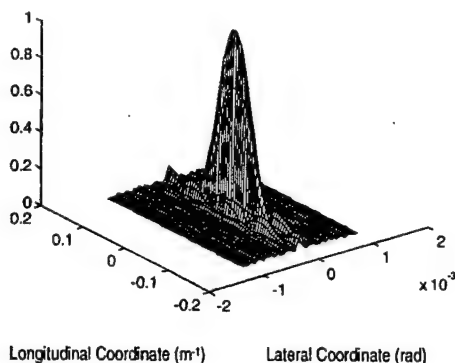
Fig. 10. Cross section of the RSI impulse response in the $x'-z'$ plane under the same constraints as for Fig. 9.

tance and the longitudinal coordinate by the square of this distance. At 2 m, for example, the center spot size is approximately 0.025 cm along the transverse axis and 14 cm along the longitudinal axis. Figure 10 is the cross section of the RSI impulse response. The values of all sampling parameters were identical for both simulations.

To compare the resolution of the two sampling schemes we use relations between lateral and longitudinal resolution and aperture size:

$$x'_{\text{res}} = \frac{C_x \lambda}{d}, \qquad z'_{\text{res}} = \frac{C_z \lambda}{d^2}, \qquad (19)$$

where x'_{res} is the lateral resolution size in radians, z'_{res} is the longitudinal resolution size in inverse length, d is the aperture diameter, and C_x and C_z are sampling-scheme-dependent unitless constants, where a smaller number indicates a smaller resolution element size or better resolution. The resolution here is not calculated with the Rayleigh two-point criterion; rather, it uses the root-mean-square size of the point-spread function (PSF). For the MSI, $C_x = 0.95$ and $C_z = 1.38$, whereas for the RSI, $C_x = 1.15$ and $C_z = 1.65$.

Under the definition of aperture used here, there is little difference in resolution between the RSI and the MSI sampling schemes. Each of the sampling schemes has its own advantages and disadvantages, however. The MSI scheme most effectively utilizes a circular aperture of a given size because the translation $(\hat{x}, \hat{y})$ is always in the same direction as the displacement $(\Delta x, \Delta y)$, so the value of q is maximized. The MSI approach will contain more correlations within a fixed sized aperture and therefore is expected to provide superior longitudinal resolution. Our simulations seem to indicate that this difference may not be great because the two methods yield similar impulse responses. The advantages of the RSI approach are that data are taken in parallel and that the higher acquisition speed makes translation of the instrument more attractive. Parallel acquisition speeds acquisition and reduces stabilization requirements. By translating the instrument as a whole one avoids the fixed-aperture assumption of our simulations, to permit a greater range for the mean

transverse displacement of sample points than for the maximum sample separation. This approach can substantially improve the resolution obtained.

4. Experimental Results

We explored our PSF models experimentally by measuring correlations produced by a laser diode with a RSI. We measured the correlations by translating the RSI laterally perpendicular to the RSI optical axis and sampling the interference intensity with various phase shifts at each lateral position.

Our RSI consisted of a 5.08-cm-aperture cube beam splitter with two 5.08-cm folding mirrors, each constructed from two separate mirrors affixed to each other at a 90-deg angle (Fig. 11). Each of the mirrors could be independently rotated about its axis such that the shear angle and the alignment axis could be set. The focal-plane array was a Princeton Instruments 512×512 backilluminated CCD camera placed at the output face of the RSI. To provide the longitudinal delay, one of the folding mirrors was placed upon a piezo-driven flexure stage, which permitted precise control of relative path length down to 10-nm resolution when it was used in conjunction with an inductive positioning sensor. All these optical components were in suitable optical mounts and bolted to a 1.9-cm-thick stainless-steel plate, which was itself bolted to a 1.27-cm-thick steel plate to provide the required vibration stability to minimize noise. The bottom plate was placed upon two steel rails, and the RSI was moved along the rails by an Aerotech translation stage, which could move the interferometer over a 5-cm distance. Even with the extremely rigid steel, the plates collectively bent enough to change the path length delay ~20 µm over its full range of travel, and this misadjustment was repeatable and corrected for when the path-length delay was set.

We measured the impulse response of this system by using a laser diode that had a center wavelength of 660 nm. The diode facets were damaged to inhibit lasing, and the device was used as a LED with a 20-nm spectral bandwidth. The source provided an elliptical radiation pattern that completely filled the aperture. An iris was used as the pupil stop to control the aperture size. The RSI imaged the source at 256 different positions of the translation stage separated by 34 µm. At each transverse RSI position, images were recorded for eight different relative path delays between the folding mirrors. These longitudinal dithers were separated by 0.125 wavelength, centered about zero path delay. The complex mutual intensity across the RSI output plane was isolated from these eight measurements in two steps. First we multiplied the 2D pattern recorded for each delay by the phase factor $\exp(-j4\pi\delta/\lambda_0)$, where δ is the path delay. Then we summed all eight modulated frames. This process isolates the component of the output intensity that oscillates at the frequency $2/\lambda_0$ under longitudinal dithering. This component is the mutual intensity. The RSI detects the mutual intensity as a function of Δx and Δy on the Cartesian
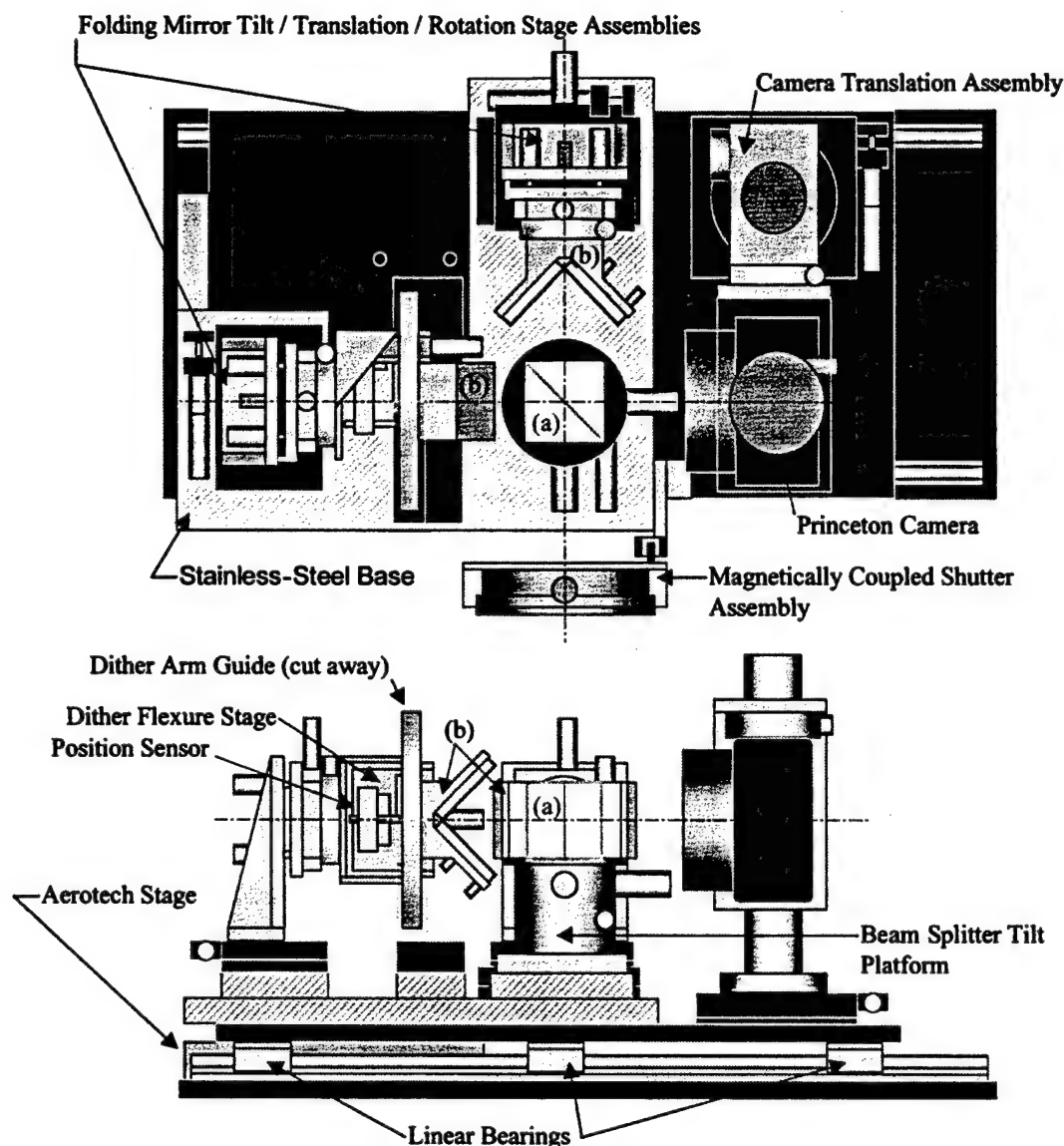
Fig. 11. RSI used to measure the mutual coherence of the four-LED test object and RSI impulse response. The RSI consisted of (a) a 5 cm × 5 cm × 5 cm cube beam splitter with (b) two 5 cm × 5 cm folding mirrors. Each folding mirror was constructed from two separate mirrors affixed to each other at a 90-deg angle, giving a full 5 cm × 5 cm square aperture. A Princeton Instruments 512 × 512 backilluminated CCD was used as the focal-plane array. For longitudinal delay, one of the folding mirrors was placed upon a piezoelectric-driven flexure stage in conjunction with an inductive positioning sensor. The RSI was mounted upon two linear bearings and was translated over a 5-cm length by an Aerotech translation stage.

CCD grid and for uniform shifts in x_g. We transformed these measurements into uniform estimates of $J_{3D}(n\Delta x, m\Delta y, l\hat{x}\Delta x)$ for integrals n, m, and l by a series of one-dimensional interpolations. We used the approximate prolate-spheroidal interpolation series[24] to implement this transformation. We then implemented a 3D fast Fourier transform of $J_{3D}(n\Delta x, m\Delta y, l\hat{x}\Delta x)$ over the indices n, l, and m to obtain the function $I(x/z, y/z, 1/z)/z^2$.

The lateral aperture diameter in our PSF experiment was 5.6 mm. The RSI was set with a 90-deg rotational shear angle, and the total translation distance was 8.7 mm to ensure full sampling of the aperture. The results of the 3D reconstructed PSF are shown in Fig. 12. Because the simulated and measured aperture sizes were so similar, there is close agreement between the sizes of the measured and the simulated PSF's. There is a slight asymmetry in the measured PSF, because the iris is not completely coincident with the axis of the RSI at the center position of the lateral travel.

We also used this experimental system to reconstruct more-complex sources. For example, we imaged a source consisting of four light-emitting diodes at $\lambda_0 = 640$ nm. In this case, no aperture stop was used to limit resolution. The mutual intensity was measured by the RSI with its shear angle set to 19 deg. The RSI was translated laterally by a micrometer-resolution translation stage to 256 different positions 193 μm apart. We obtained a 2D measure of the
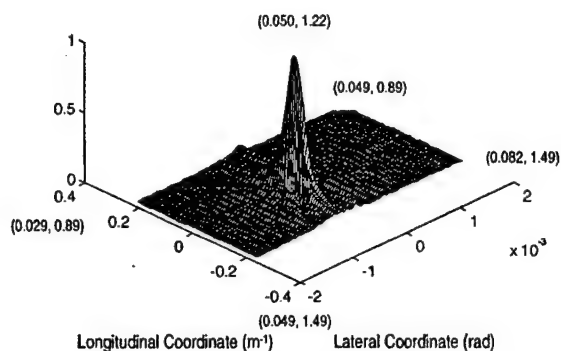
Fig. 12. Experimental cross section of the RSI impulse response in the x'–z' plane. The four corners of the plane and the peak are labeled with their Cartesian coordinates in real space, in meters, relative to the origin of the focal-plane array. This impulse response was sampled by a linearly translated RSI by use of the procedure and the experimental parameters described in the text.
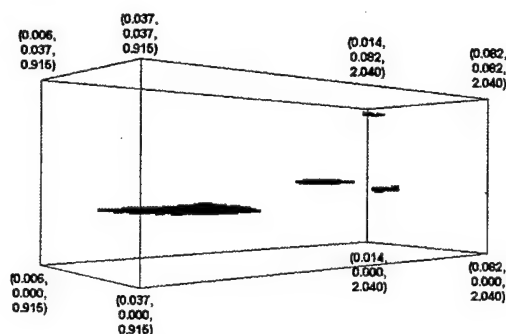

Fig. 13. Experimental reconstruction of a four-light-emitting-diode test source, as sampled by the RSI. The 50% power density isosurface is shown. The LED's appear to be of different sizes because they were in fact of different intensities. The source is shown in projective coordinates, but the corners are labeled in Cartesian coordinates, in meters, relative to the origin of the focal-plane array. These data were taken by a RSI by use of the procedure and the experimental parameters described in the text.

mutual intensity at each position, using the eight-longitudinal-position approach described above for the PSF measurement. Figure 13 shows the estimated power density of the LED sources as a 50% constant isosurface of the maximum power density in the source. Because the LED's provide differing intensities, each appears to be a different size, when they were in fact all similarly sized. Three of the LED's were in a rear plane approximately 1.5 m from the RSI pupil plane and one was 1 m away. The longitudinal accuracy of reconstruction was approximately 0.2 m^{-1}, or 20 cm at a 1-m distance. These results demonstrate that lensless imaging of 3D sources with coherence measurements alone is possible.

5. Conclusion

We have shown that finite-aperture 3D coherence imaging can be extended to the Fresnel diffraction zone by straightforward Fourier analysis, and we have analyzed the resolution of both Fraunhofer- and

Fresnel-zone imaging. The PSF of a rotational shearing interferometer was also experimentally measured. Although coherence imaging holds the potential to revolutionize 3D imaging, several further questions remain to be addressed. Most notably, noise issues are not addressed here but will play an important role in computational coherence imaging. It is interesting to note that, in contrast with the point-to-point independence of conventional imaging systems, system noise scales with object complexity in coherence imaging systems. When the information capacity of a conventional imaging system is limited only by the space–bandwidth product, the information capacity of a coherence imaging system will be limited by both the space–band volume product and noise scaling.

References

1. C. V. Schooneveld, ed., *Image Formation from Coherence Functions in Astronomy*, Vol. 76 of International Astronomical Union Colloquium 49 (Reidel, Dordrecht, The Netherlands, 1978).
2. F. Roddier, "Interferometric imaging in optical astronomy," Phys. Rep. **170**, 97–166 (1988).
3. G. W. Swenson, "Radio astronomy precedent for optical interferometer imaging," J. Opt. Soc. Am. A **3**, 1311–1319 (1986).
4. J. T. Armstrong, D. J. Hutter, K. J. Johnston, and D. Mozurkewich, "Stellar optical interferometry in the 1990s," Phys. Today **48**(5), 42–49 (1995).
5. A. J. Devaney, "The inverse problem for random sources," J. Math. Phys. **20**, 1687–1691 (1979).
6. W. H. Carter and E. Wolf, "Correlation theory of wavefields generated by fluctuating, three-dimensional, primary, scalar sources. I. General theory," Opt. Acta **28**, 227–244 (1981).
7. I. J. LaHaie, "Inverse source problem for three-dimensional partially coherent sources and fields," J. Opt. Soc. Am. A **2**, 35–45 (1985).
8. A. M. Zarubin, "Three-dimensional generalization of the van Cittert–Zernike theorem to wave and particle scattering," Opt. Commun. **100**, 491–507 (1993).
9. J. Rosen and A. Yariv, "General theorem of spatial coherence: application to three-dimensional imaging," J. Opt. Soc. Am. **13**, 2091–2095 (1996).
10. J. Rosen and A. Yariv, "Reconstruction of longitudinal distributed incoherent sources," Opt. Lett. **21**, 1803–1806 (1996).
11. J. Rosen and A. Yariv, "Three-dimensional imaging of random radiation sources," Opt. Lett. **21**, 1011–1014 (1996).
12. B. R. Frieden, "Optical transfer of the three-dimensional object," J. Opt. Soc. Am. **57**, 56–66 (1967).
13. A. W. Lohmann, "Three-dimensional properties of wavefields," Optik **51**, 105–117 (1978).
14. L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics* (Cambridge U. Press, Cambridge, 1995).
15. M. Y. Chiu, H. H. Barrett, R. G. Simpson, C. Chou, J. W. Ardent, and G. R. Gindi, "Three-dimensional radiographic imaging with a restricted view angle," J. Opt. Soc. Am. **69**, 1323–1333 (1979).

16. D. H. DeVorkin, "Michelson and the problem of stellar diameters," J. Hist. Astron. **6**, 1–18 (1975).

17. J. D. Armitage and A. Lohmann, "Rotary shearing interferometry," Opt. Acta **12**, 185–192 (1965).

18. C. Roddier and F. Roddier, "Imaging with a coherence interferometer in optical astronomy," in *Image Formation from Coherence Functions in Astronomy*, C. V. Schooneveld, ed., Vol. 76 of International Astronomical Union Colloquium 49 (Reidel, Dordrecht, The Netherlands, 1979), pp. 175–179.

19. K. Itoh and Y. Ohtsuka, "Fourier-transform spectral imaging: retrieval of source information from three-dimensional spatial coherence," J. Opt. Soc. Am. A **3**, 94–100 (1986).

20. K. Itoh, T. Inoue, T. Yoshida, and Y. Ichioka, "Interferometric supermultispectral imaging," Appl. Opt. **29**, 1625–1630 (1990).

21. K. Itoh, T. Inoue, and Y. Ichioka, "Interferometric spectral imaging and optical three-dimensional Fourier transformation," J. J. Appl. Phys. **29**, L1561–L1564 (1990).

22. K. Itoh, "Interferometric multispectral imaging," in *Progress in Optics*, E. Wolf, ed. (North-Holland, Amsterdam, 1996), Vol. 35, pp. 145–196.

23. L. Mertz, *Transformations in Optics* (Wiley, New York, 1965).

24. J. J. Knab, "Interpolation of band-limited functions using the approximate prolate series," IEEE Trans. Inf. Theory **IT-25**, 717–720 (1979).

Astigmatic Coherence Sensors

Astigmatic coherence sensor for digital imaging

Daniel M. Marks, Ronald A. Stack, and David J. Brady

*Beckman Institute for Science and Technology and Department of Electrical and Computer Engineering,
University of Illinois at Urbana-Champaign, 405 N. Mathews Avenue, Urbana, Illinois 61801*

We present a novel sensor that measures the entire spatial coherence function within an aperture by use of
a variable astigmatic lens. This sensor permits digital capture and processing of partially coherent fields.
We demonstrate the sensor by sampling and computing the coherent modes of a three-dimensional incoherent
source.   © 2000 Optical Society of America
  OCIS codes: 110.1650, 030.6600.

With digital sensing and processing becoming commonplace in optical imaging, one wonders if a nonimaging sensor may allow a wider range of information about the optical field to be sampled and processed. For example, the full set of correlations between points within a two-dimensional aperture forms a four-dimensional (4-D) set of potentially independent data. If the entire coherence function in the aperture could be sampled, a new analysis of the data would be available that cannot be performed on incomplete data. In this Letter we propose a new type of sensor, the astigmatic coherence sensor (ACS), which is able to sample the 4-D coherence function in an aperture. The ACS addresses the stability and signal-to-noise limitations associated with other coherence-sensing instruments such as the Michelson stellar interferometer[1,2] and the rotational shearing interferometer.[3,4] Mechanical stability problems can make sampling large amounts of coherence data difficult, and detecting weak signals in noise requires long integration periods. The method used by the ACS is related to other techniques such as depth-by-defocus imaging[5,6] and phase diversity imaging.[7] Furthermore, interpretation of data from the sensor provides new insight into the sampling of coherence in standard imaging systems.

The ACS consists of a nonspherical lens, which has independently changeable horizontal and vertical focal lengths f_x and f_y, respectively. Although such a lens is difficult to realize in practice, here we introduce a practical substitute. This lens is placed a distance z from a focal-plane array such as a CCD, and the horizontal and vertical locations on the CCD relative to the center axis of the lens are positions x and y, respectively. A partially coherent, quasi-monochromatic field with mutual intensity $J(x_1, y_1, x_2, y_2)$ is incident on the aperture of the lens. The intensity measured by the focal-plane array is then

$$I(x, y, z, f_x, f_y) = \frac{1}{z^2} \int_A \int_A J(x_1, y_1, x_2, y_2)$$

$$\times \exp\left\{ i \frac{2\pi}{\lambda} \left[\left(\frac{x_1{}^2}{f_x} + \frac{y_1{}^2}{f_y} \right) - \left(\frac{x_2{}^2}{f_x} + \frac{y_2{}^2}{f_y} \right) \right] \right\}$$

$$\times \exp\left\{ \frac{-i2\pi}{\lambda z} \left[(x - x_1)^2 + (y - y_1)^2 \right.\right.$$

$$\left.\left. - (x - x_2)^2 - (y - y_2)^2 \right] \right\} dx_1 dy_1 dx_2 dy_2 . \quad (1)$$

If we make the substitutions $x_1 = \hat{x} + \Delta x$, $x_2 = \hat{x} - \Delta x$, $y_1 = \hat{y} + \Delta y$, and $y_2 = \hat{y} - \Delta y$, the integral becomes

$$z^2 I(x, y, z, f_x, f_y) = \frac{1}{2} \int_A \int_A J(\Delta x, \Delta y, \hat{x}, \hat{y})$$

$$\times \exp\left\{ i \frac{2\pi}{\lambda} \left[4\hat{x}\Delta x \left(\frac{1}{f_x} - \frac{1}{z} \right) + 4\hat{y}\Delta y \left(\frac{1}{f_y} - \frac{1}{z} \right) \right] \right\}$$

$$\times \exp\left[\frac{-i2\pi}{\lambda z} (-4x\Delta x - 4y\Delta y) \right] d\hat{x} d\hat{y} d\Delta x d\Delta y . \quad (2)$$

We make another set of substitutions, $q_x = \hat{x}\Delta x$, $dq_x/|\Delta x| = d\hat{x}$, $q_y = \hat{y}\Delta y$, and $dq_y/|\Delta y| = d\hat{y}$, to find

$$z^2 I\left(\frac{x}{z}, \frac{y}{z}, \frac{1}{f_x} - \frac{1}{z}, \frac{1}{f_y} - \frac{1}{z} \right) = \frac{1}{2} \int_A \int_A$$

$$\times \left(\begin{array}{c} \dfrac{J(\Delta x, \Delta y, q_x, q_y)}{|\Delta x \Delta y|} \\[2mm] \exp\left\{ i \dfrac{2\pi}{\lambda} \left[4q_x \left(\dfrac{1}{f_x} - \dfrac{1}{z} \right) + 4q_y \left(\dfrac{1}{f_y} - \dfrac{1}{z} \right) \right] \right\} \\[2mm] \exp\left[\dfrac{-i2\pi}{\lambda z} (-4x\Delta x - 4y\Delta y) \right] \end{array} \right)$$

$$\times dq_x dq_y d\Delta x d\Delta y . \quad (3)$$

The intensity-measurement function $I(\cdot)$ of Eq. (3) contains the same measurements as in Eq. (2) resampled to new coordinates. Examination of Eq. (3) reveals that there is a 4-D Fourier-transform relationship between the following quantities:

$$z^2 I\left(\frac{x}{z}, \frac{y}{z}, \frac{1}{f_x} - \frac{1}{z}, \frac{1}{f_y} - \frac{1}{z} \right) \Leftrightarrow$$

$$J\left(\frac{\Delta x}{\lambda}, \frac{\Delta y}{\lambda}, \frac{q_x}{\lambda}, \frac{q_y}{\lambda} \right) \Big/ |\Delta x \Delta y| . \quad (4)$$

Every sample of intensity for each combination of the position and the focal lengths of the lens is a sample of the 4-D Fourier transform of the coherence function. This result applies equally well for $f_x = f_y$, so stigmatic imagers also measure samples of the 4-D Fourier transform of the coherence function, but they are unable to sample the entire Fourier space.

Because the ACS can measure general partially coherent sources, it can be used equally well with coherent or incoherent sources and can distinguish between

the two. This ability comes at the expense of the requirement of much more information than would be the case if the coherence state of the source were known. Another advantage of measuring the entire coherence function is that a coherence-mode decomposition [8,9] can be performed. This decomposition will allow the contributions of individual sources to be separated. Without the entire coherence function, we will need to know the coherence state of the sources *a priori* to separate their contributions. The coherence-mode decomposition can become a powerful computational technique for augmenting the imaging process. In addition, if the entire coherence function is known, the coherence can be found after propagation through any linear optical system, including any other optical instrument. A new kind of image processing is possible in which the propagation of partially coherent light can be digitally simulated.

An ACS might be constructed by use of two cylindrical lenses of focal length f with their focal axes placed at an angle θ relative to each other, symmetric about the horizontal axis. The effective horizontal and vertical focal lengths of this lens are given by $2\cos^2(\theta/2)/f = 1/f_x$ and $2\sin^2(\theta/2)/f = 1/f_y$, respectively. To achieve any given values of $1/f_x - 1/z$ and $1/f_y - 1/z$, one need only set θ and the distance between the lens combination and the CCD z.

Because the aperture is a finite size, the accessible region of the 4-D coherence function will be limited. If we consider a square aperture of side d, we can determine the accessible region of the coherence space. To keep the correlations confined within the aperture, we require that $|\hat{x} \pm \Delta x| < d$ and $|\hat{y} \pm \Delta y| < d$. A plot of the boundaries denoted by these inequalities is given in Fig. 1. Because the Fourier-space parameters are different from the physical parameters, the region of the Δx and q_x Fourier space that these boundaries correspond to is also shown in Fig. 1. Since the Fourier space is not sampled near the q_x axis, an image formed by a finite-aperture sensor will suffer in resolution along this dimension.

We built an ACS using three lenses: one 15-cm cylindrical lens used for one axis and two 30-cm cylindrical lenses used for the other axis. A diagram of this scheme is shown in Fig. 2. Two lenses were used for one axis instead of one lens as we described above so that the principal planes of focus on both axes could roughly coincide. The two lenses for one axis were always turned together. There were two computer-controlled rotation stages, each turning one axis of the ACS. The CCD was placed on a computer-controlled translation stage so that the distance between the lens system and the CCD could be changed. The translation stage had a travel of 5 cm, with its center position being 15 cm from the principal planes of the lenses. The source that we imaged with the ACS was three red LED's with their plastic lenses sanded off to make their radiation patterns more isotropic.

The data acquisition went as follows: The computer stepped through 64 values of $1/f_x - 1/z$ and $1/f_y - 1/z$ spaced evenly by 0.0035 m^{-1} for a total range of 2.2 m^{-1}. For each of the pairs of values

the computer calculated the angle θ and distance z needed to achieve these values. If the position could be reached by the translation stage, the computer set the rotation stages at equal and opposite angles θ about the horizontal axis and set the position of the stage. The computer then sampled the intensity on a 256×256 region of the CCD array, which is downsampled by use of a band-limiting interpolator to 64×64. For roughly one quarter of the images, the position could not be reached by the translation stage, so the computer recorded zeros for the picture and continued sampling. We recorded a total of 4096 64×64 pixel frames in this way in a total time of ~12 h, to record a total of 2^{24} samples.

After the data were recorded, they were processed to find the coherent modes. First, we performed a 4-D fast Fourier transform of dimensions $64 \times 64 \times 64 \times 64$ on the data to compute the discrete samples of the coherence function from the sampled intensity. Then, the Lanczos method was used to calculate the approximate eigenvalues and eigenvectors of the coherence data, which correspond to the coherent modes. The algorithm was adapted for this purpose by approximation of the integral that defines the coherent-mode eigenvalue equation with a discrete sum:
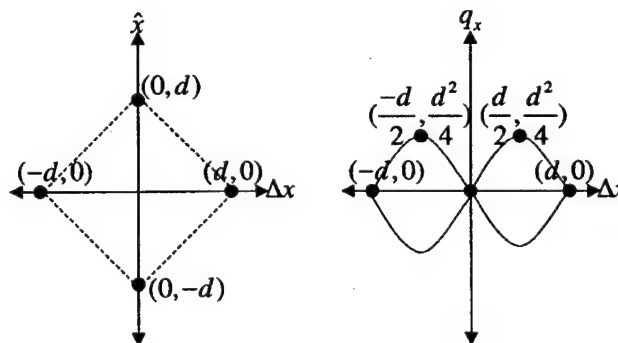
Fig. 1. Portion of coherence space that can be sampled in a two-dimensional square aperture (shown for one dimension only; the other is identical).
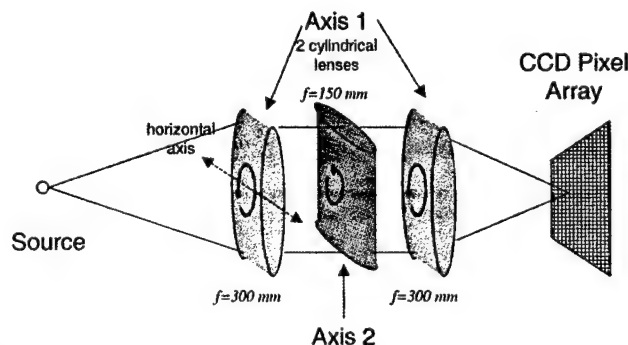
Fig. 2. Diagram of the ACS. Two cylindrical lenses of 300-mm focal length form Axis 1 and focus along one diagonal direction, and one cylindrical lens of 150-mm focal length forms Axis 2. The combination of the two foci, oriented at equal and opposite angles to the vertical axis, effectively forms a single lens of adjustable astigmatic focal ratio.
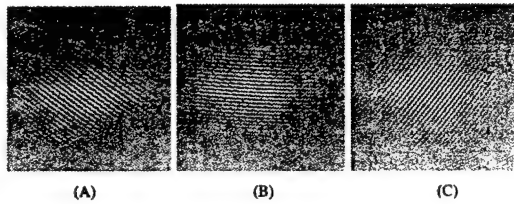
Mode	Source magnitude (arb units)	X spatial frequency (m⁻¹)	Y spatial frequency (m⁻¹)	Curvature of field (m⁻²)	Position X (m)	Position Y (m)	Distance Z (m)
A	8.86×10^5	10400	-7500	3.98×10^5	0.0050	-0.0036	0.758
B	8.64×10^5	-15300	700	6.16×10^5	-0.0048	.000216	0.496
C	7.43×10^5	8900	11400	3.98×10^5	0.0042	0.0055	0.758

Fig. 3. (A)–(C) Three coherent modes computed from the sample coherence data. The image intensity corresponds to the magnitude of the real component of the unpolarized optical field. The dimensions of the images are 3 mm × 3 mm. The magnitude, spatial frequency, field curvature, and corresponding positions of each mode is shown in the table.

$$E(\mathbf{r}_2) = \lambda_k \int_A E(\mathbf{r}_1) J(\mathbf{r}_1, \mathbf{r}_2) d\mathbf{r}_1 \rightarrow$$

$$E_j = \lambda_k \sum_i E_i J\left(\mathbf{r}_i - \mathbf{r}_j, \frac{\mathbf{r}_i + \mathbf{r}_j}{2}\right), \qquad (5)$$

where $\mathbf{r}_i$ are the positions of the points of interest in the aperture (a rectangular array here), λ_i are the eigenvalues, and $J(\cdot)$ is a function of the center and difference positions of the correlated points. For the points at which no data for $J(\cdot)$ were gathered owing to the limited measurement range of the sensor, we set $J(\cdot) = 0$. We expect that the incomplete data will artificially increase the number of coherent modes and result in a coherence function that is no longer perfectly positive definite.

Figure 3 shows the coherent modes factored from the cross-spectral density by the Lanczos method.[10] The primary modes are all spherical waves, but there were higher-order modes present in the signal because the sources were not perfectly pointlike. The spatial-frequency (plane-wave) component of these spherical waves corresponds to their angular coordinates, whereas the curvature of the wave front corresponds to the depth. By fitting a Fresnel wave front to each mode, we obtain

$$\exp\left\{\frac{2\pi i}{\lambda}\left[\frac{x}{z}n_x d + \frac{y}{z}n_y d + \frac{1}{2z}(n_x^2 + n_y^2)d^2\right]\right\},$$

where x, y, and z are the position of the source; n_x and n_y are the pixel numbers in the field; and $d = 2.35 \times 10^{-5}$ m is the size of the resolution element with which the field was measured. The lateral distance between the first and third LED's was 9 mm as measured with a ruler and 9.1 mm as measured with the coherence sensor. The distance in depth between the second LED and the other two was 245 mm as measured with a ruler and 262 mm as measured with the sensor. There was good agreement between these two methods, showing that the sensor can measure coherent modes accurately.

We have proposed and demonstrated a sensor that can sample a partially coherent field and uses digital processing in the form of a discretized coherent-mode transformation to identify individual sources in the field. We believe that such sensors will be useful not only for three-dimensional (3-D) incoherent sources such as those that were imaged here but also for more-general 4-D coherence functions in which distortions break the symmetry associated with 3-D coherence propagation.[11] Because in the ACS every source does not contribute light equally to every pixel measured, as in other methods of sampling 3-D structure from coherence,[12,13] such as the rotational shearing interferometer, the ACS may have the advantage of reduced noise compared with that of white-light holography.[14] Also, because the ACS does not employ two separate interferometer arms, it is less sensitive to relative motions between its components.

References

1. A. A. Michelson, Philos. Mag. **30,** 1 (1890).
2. A. Michelson and F. G. Pease, Astrophys. J. **53,** 249 (1921).
3. M. Murty, J. Opt. Soc. Am. **54,** 1187 (1964).
4. F. Roddier, in *High Angular Resolution Stellar Interferometry*, J. Davis and W. J. Tango, eds., Vol. 50 of IAU Colloquia (University of Sydney, Sydney, Australia, 1979), paper 3.
5. A. Pentland, S. Scherock, T. Darrell, and B. Girod, J. Opt. Soc. Am. A **11,** 2925 (1994).
6. S. K. Nayar, M. Watanabe, and M. Noguchi, IEEE Trans. Pattern Anal. Mach. Intell. **18,** 1186 (1996).
7. R. G. Paxman, T. J. Schulz, and J. R. Fienup, J. Opt. Soc. Am. A **9,** 1072 (1992).
8. E. Wolf, J. Opt. Soc. Am. **72,** 343 (1982).
9. L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics* (Cambridge U. Press, Cambridge, England, 1995).
10. G. H. Golub and C. F. Van Loan, *Matrix Computations* (Johns Hopkins U. Press, Baltimore, Md., 1996).
11. D. L. Marks, R. A. Stack, and D. J. Brady, Appl. Opt. **38,** 1332 (1999).
12. D. L. Marks, R. A. Stack, D. J. Brady, D. Munson, and R. B. Brady, Science **284,** 2164 (1999).
13. J. Rosen and A. Yariv, Opt. Lett. **21,** 1803 (1996).
14. E. Ribak, C. Roddier, F. Roddier, and J. Breckinridge, Appl. Opt. **27,** 1183 (1988).

Digital Refraction Distortion Correction using an Astigmatic Coherence Sensor

D. L. Marks, R. A. Stack, and D. J. Brady

Beckman Institute and Electrical and Computer Engineering Department

University of Illinois at Urbana-Champaign, 405 N. Mathews, Urbana IL 61801

Distorted wavefronts have been characterized using interferometry [1], holography [2], phase diversity [3, 4], modal analysis [5], and wavefront sensors [6]. A distorted wavefront can be corrected [7, 8] using adaptive optics [9], liquid crystal phase modulators [10, 11, 12], or digital deconvolution from wavefront sensing [13, 14]. We describe here a method of digital wavefront sensing and deconvolution based on the recently described Astigmatic Coherence Sensor (ACS) [15]. The ACS measures the four-dimensional spatial coherence function within an aperture. In this paper we use the ACS to digitally characterize and correct a wavefront distortion in an imaging system. This method is unique in that it employs an extremely powerful analytic tool for analyzing partially coherent sources, the coherent-mode decomposition [17]. This method can only be employed when all the correlations are measured between a set of points, as the ACS does.

The ACS consists of a lens combination that has an adjustable horizontal to vertical focal length ratio. A sensor array placed behind the lenses and measures the intensity of the field. The intensity $I(x, y, z)$ is measured in the x-y plane transverse to the optical axis and as a function of z, the distance from the sensor array to the principal plane of the lens combination.

1

$I(x, y, z)$ is related to the mutual intensity of the quasi monochromatic field of wavelength λ in the aperture, $J(\hat{x} - \Delta x, \hat{y} - \Delta y, \hat{x} + \Delta x, \hat{y} + \Delta y)$, by a four-dimensional Fourier transform:

$$z^2 I \left(\frac{x}{z}, \frac{y}{z}, \frac{1}{f_x} - \frac{1}{z}, \frac{1}{f_y} - \frac{1}{z} \right) =$$

$$\frac{1}{2} \int_A \int_A \left(\begin{array}{c} \frac{J(\Delta x, \Delta y, q_x, q_y)}{|\Delta x \Delta y|} \\ \exp \left(i \frac{2\pi}{\lambda} \left[4 q_x \left(\frac{1}{f_x} - \frac{1}{z} \right) + 4 q_y \left(\frac{1}{f_y} - \frac{1}{z} \right) \right] \right) \\ \exp \left(\frac{-i2\pi}{\lambda z} \left[-4 x \Delta x - 4 y \Delta y \right] \right) \end{array} \right) dq_x dq_y d\Delta x d\Delta y \qquad (1)$$

where f_x and f_y are the focal lengths of the astigmatic lens combination along the x and y axes. [15] For an incoherent source, the domain of $J(\cdot)$ reduces to three dimensions, because incoherent sources produce only independent spherical waves. However, when incoherent sources are imaged through a distortion, the distorting medium can break the symmetry of spherical waves, and produce independent data in a 4-D domain. In some cases knowledge of the 4-D mutual intensity can be used to recover an undistorted image of the original source. In this paper we use measurements of $J(\cdot)$ in 4-D to compensate for an isoplanatic (inside the pupil) phase distortion.

The coherent-mode decomposition expresses a partially coherent field as an orthogonal expansion of analytic wavefronts from uncorrelated sources: $J(\mathbf{r_1}, \mathbf{r_2}) = \sum_i \lambda_i \phi_i(\mathbf{r_1}) \phi_i^*(\mathbf{r_2})$, where the eigenvalues λ_i correspond to the power emitted from each source, and the orthogonal eigenfunctions $\phi_i(\mathbf{r})$

2

are the complex wavefronts from each source in the entrance pupil S. We consider the transmission of a partially coherent wavefront through an optical system from an entrance pupil S to an exit pupil S'. The optical system has a coherent transfer function $H(\mathbf{r_1}, \mathbf{r_1'}) = \sum_j S_j \psi_j^*(\mathbf{r_1})\psi_j'(\mathbf{r_1'})$, where the S_j are positive singular values and the $\psi_j(\mathbf{r_1})$ and $\psi_j'(\mathbf{r_1'})$ are the orthogonal singular functions in S and S', normalized to one. The functions $\phi_i'(\mathbf{r_1'}) = \sum_j S_j \int_S \phi_i(\mathbf{r})\psi_j^*(\mathbf{r}) \, d\mathbf{r}$ are the wavefronts in the exit pupil of the entrance function for each coherent mode. We can then calculate the overlap matrix $\mathbf{P}_{ij} = \int_{S'} \phi_i'^*(\mathbf{r_1'})\phi_j(\mathbf{r_2'}) = \sum_k \mathbf{A}_{ik}^* S_k^2 \mathbf{A}_{jk} = \mathbf{A}^\dagger \mathbf{S}^2 \mathbf{A}$ (in matrix notation), where $\mathbf{A}_{ij} = \int_S \phi_i(\mathbf{r})\psi_j^*(\mathbf{r}) \, d\mathbf{r}$ and $\mathbf{S}_{ij}^2 = S_j^2 \delta_{ij}$.

The matrix $\mathbf{P}$ is determined by how orthogonal the coherent modes are after they pass through the system. If the functions $\phi_i(\mathbf{r})$ and $\psi_i(\mathbf{r})$ form a complete set within S, then the matrix $\mathbf{A}$ is unitary, and the decomposition $\mathbf{P} = \mathbf{A}^\dagger \mathbf{S}^2 \mathbf{A}$ is a eigenvalue decomposition of $\mathbf{P}$ with eigenvalues given by $\mathbf{S}^2$. Therefore, they have the same eigenvalues, determinant, and trace. The values S_j^2 correspond to the fraction of power in the entrance wavefront that projects on to the singular function $\psi_j(\mathbf{r})$ that exits the pupil. If no power is lost, the $S_j = 1$ and the $\mathbf{P} = \mathbf{I}$, the identity matrix, so all the coherent modes are still orthogonal upon leaving the exit pupil. However, loss of power in general leads to a loss of orthogonality among the coherent modes. As a result, sources that are uncorrelated appear partially correlated in the aperture, due to a loss of information. The coherent modes at the exit pupil

3

will no longer correspond one-to-one with input modes. This can be seen when looking at two closely spaced point sources through a aperture too small to resolve them. For a phase distortion placed at the exit pupil, no power can be scattered out of the optical system and so orthogonality is preserved. More general distortions such as anisoplanatic distortions will in general scatter light away from the exit pupil, and the coherent modes from which the most light is diverted will probably be "mixed" the most with other modes.

As long as one knows that a given optical system is power-preserving, a coherence-mode decomposition will separate the wavefronts due to independent sources *even without knowledge* of the exact transformation between the source and exit pupil. By examining the wavefronts due to several independent sources, one may be able to infer information about the intervening optical system. For example, if a phase distortion is placed in the pupil of an optical system viewing a planar incoherent source, all of the coherent modes upon exit will have the same planar distortion on them. From the coherent modes, one may be able to simultaneously deduce the distortion and the source behind it, even in situations when the sources can not be imaged separately or turned on/off in sequence. When the pupil is too small, or intervening blockages or distortions absorb or divert power, the coherent modes can be expected to mix in a way that is varies continuously with the amount of lost power from the source. Quantifying this "mixing" of the

4

coherent modes based on the type of distortion is beyond the scope of this paper.

Our goal is to image a test object through an unknown distortion. As illustrated in Figure 1, we use the ACS to measure the mutual intensity due to the test object and a point reference source propagation through a thin distorter. Both sources are quasi-monochromatic with the same nominal wavelength. The mutual intensity $J(x_1, y_1, x_2, y_2)$ of the combined source is the sum of the mutual intensity due to the point reference and the test object. To separate the contributions of the two, we use the coherent mode decomposition. [16, 17] The coherent mode decomposition expands the mutual intensity as a sum of coherent fields multiplied by uncorrelated random variables: $J(x_1, y_1, x_2, y_2) = \sum_i \hat{\lambda}_i \phi_i(x_1, y_1) \phi_i^*(x_2, y_2)$. Since the fields produced by the point source and illuminated transparency will be uncorrelated, we can apply the coherent mode decomposition to the mutual intensity to find the field due to the point source alone.

We assume that the distortion is isoplanatic with a pupil function $T(x, y)$ in the aperture. This distortion transforms the mutual intensity into $J_D(x_1, y_1, x_2, y_2) = J(x_1, y_1, x_2, y_2) T(x_1, y_1) T^*(x_2, y_2)$. Since the point source would produce a spherical wave coherent mode absent the distortion, we can use the actual coherent mode of the point source to characterize the distortion $T(x, y)$. After performing the coherent mode decomposition on $J'(\cdot)$, we find the coherent mode $\phi_1(x, y)$ corresponding to the distorted point source and form the con-

5

jugate phase distortion $\phi'_1(x, y)$. We then find an estimate of the coherence function before distortion $J'(x_1, y_1, x_2, y_2) = J_D(x_1, y_1, x_2, y_2)\phi'_1(x_1, y_1)\phi'^*_1(x_2, y_2)$. This estimate has the distortion corrected, and also images the transparency, because $\phi'_1(x, y)$ also conjugates the phase due to propagation of the field. We then use a 4-D inverse Fourier Transform to recover the intensity an imaging system would have measured at the in focus plane, which we expect will be the undistorted image of the source.

Our implementation of the ACS uses three lenses. Two of the lenses are cylindrical are of 300 mm focal length with their focal axes aligned and rotated to an angle $\theta/2$ from the horizontal axis. The third lens is a 150 mm focal length cylindrical lens and is placed between the other two with its axis placed $-\theta/2$ from the horizontal axis. Together they form a lens with horizontal focal length $1/f_x = 2\cos^2(\theta/2)/f$ and $1/f_y = 2\sin^2(\theta/2)/f$ vertical focal length, where $f = 150$ mm. The reference point source in our experiments was a 4 mW 660 nm laser diode attenuated by three neutral density filters with a total optical density of 4.6. Unlike a normal reference point source as used in adaptive optics, our source was not a separable wavelength. Rather, it was incorporated to include an object in the scene of sufficiently high spatial bandwidth to reconstruct the distortion. The test object was a laser printer transparency made diffuse by rubbing the back surface with sandpaper, with transparent letters "UI" surrounded by a black opaque background. This object was rear illuminated by seven red LEDs.

6

The light from the test object and the point reference was combined using a beam splitter. The objects were approximately 30 cm from the aperture of the ACS. The distortion plate was an approximately 5 by 5 cm square of 2 mm thick transparent acrylic, which was softened by heating and twisted into a distorting shape. It was placed about 15 cm in front of the ACS to make the source not resolvable by a standard stigmatic imager. Figure 1 shows a diagram of the source and the ACS. Figure 2 shows a picture of the source taken through the distortion before correction. We note that with the distortion, the images of the laser diode and transparency are inseparable.

The pixel pitch of the CCD sensor was $p = 19\mu m$, and it was nominally located $l = 185$ mm from the principal plane of the ACS. To acquire intensity data, θ and the CCD range were adjusted so that the defocus parameters $1/f_x - 1/z$ and $1/f_y - 1/z$ were stepped through combinations of 64 positions 0.034 m^{-1} apart to sample a total range of 2.13 m^{-1}. At each defocus setting, a 64x64 region of the CCD was sampled. The size of the sampled region was proportional to the distance z away from the principal plane. A bandlimiting interpolator resampled the pixels to have a spacing slightly above or below $19\mu m$ as needed. The resolution of the measured coherence was $d = \frac{\lambda l}{Np} = 96$ μm. Ultimately 4096 images were recorded in 12 hours.

We recovered a deblurred image from the coherence data by the following steps:

1. The data from the ACS was measured as a 64x64x64x64 discretized

version of $I(x/z, y/z, 1/f_x - 1/z, 1/f_y - 1/z)$. The sampling rate in the x/z and y/z variables was $1.02 \ 10^{-4}$, and the sampling rate in $1/f_x - 1/z$ and $1/f_y - 1/z$ was $0.034 \ m^{-1}$. Using a radix-2 4-D real-to-complex FFT this was converted to the discrete coherence samples $J(\Delta x, \Delta y, \hat{x}\Delta x, \hat{y}\Delta y)/|\Delta x \Delta y|$. This data set was only 32x64x64x64 because it was the Fourier transform of a real function.

2. We then multiplied the 4-D coherence by a filter $|\Delta x \Delta y|$ to eliminate the $\Delta x \Delta y$ factor in Eqn. 1. To avoid the loss of this data for $J(\cdot)$ such that $\Delta x = 0$ or $\Delta y = 0$, we multiplied by a small number (0.5) instead of zero.

3. We used the Lanczos sparse matrix eigenvalue algorithm [18] find the coherent mode decomposition of $J(\cdot)$. This was done by discretizing the integral that defines the coherent mode decomposition:

$$\phi\left(\mathbf{r_2}\right) = \lambda_k \int_A \phi\left(\mathbf{r_1}\right) J\left(\mathbf{r_1}, \mathbf{r_2}\right) d\mathbf{r_1} \rightarrow \phi_j = \lambda_k \sum_i \phi_i J\left(\mathbf{r}_i - \mathbf{r}_{j,} \frac{\mathbf{r}_i + \mathbf{r}_j}{2}\right)$$

(2)

where $\mathbf{r}_i$ are the positions of the points of interest in the aperture, the λ_i are the eigenvalues, and $J\left(\cdot\right)$ is a function of the center and difference positions of the correlated points. In the Lanczos algorithm, the sampling points of $J(\cdot)$ and $\phi(\cdot)$ do not coincide because they are measured on separate coordinate systems. To perform the matrix-vector multiply step in the algorithm, a quadrilinear interpolator was

8

used to find samples of $J(\cdot)$ from the 16 nearest samples.

4. We found the 64x64 sampled field corresponding to the highest eigen-value coherent mode $\phi_1(x, y)$, which was due to the laser diode. This field is sampled with a period of 96 μm. The real part of this analytic field is shown in Figure 3, and the image dimensions are 6.1x6.1 mm.

5. We computed an inverse filter $\phi'(x, y)$ from this using the following:

$$\phi'(x, y) = \frac{\phi_1^*(x, y)}{|\phi_1(x, y)| + 0.01\phi_{max}} \tag{3}$$

This essentially produced a filter with a conjugate phase to the distortion. The ϕ_{max} term is the maximum magnitude of the field within the aperture and was added to suppress noisy field points with low magnitudes. This is a noise reduction approach similar to that of a Wiener filter.

6. We formed a nonblurred sampled coherence function:

$J'(\Delta x, \Delta y, \hat{x}\Delta x, \hat{y}\Delta y) =$

$J(\Delta x, \Delta y, \hat{x}\Delta x, \hat{y}\Delta y)\phi'(\hat{x} - \Delta x, \hat{y} - \Delta y)\phi'^*(\hat{x} + \Delta x, \hat{y} + \Delta y)/|\Delta x \Delta y|$

$$\tag{4}$$

The denominator was added to reverse step 2 so we can convert the result back to intensity data. When Δx or Δy is zero, we multiply the sample by 2 to reverse the 0.5 factor of step 2 for these points. Since, as in step 3, the samples of $J(\cdot)$ and $\phi'(\cdot)$ do not coincide, a bilinear

9

interpolation of $\phi'(\cdot)$ is performed between the nearest 4 neighbors of a needed point.

7. We used a 4-D inverse complex-to-real FFT to reverse step 1 and recover $I(x/z, y/z, 1/f_x - 1/z, 1/f_y - 1/z)$

8. Finally, the plane of data corresponding to the in focus image was extracted, and is displayed in Figure 4.

We believe that the legibility of the letters has been substantially improved. Also note that the reference diode is pointlike, meaning the blurring of the laser diode has been successfully removed. The image of the diode is not a single pixel because the bandwidth of the imaging system was too small to image it as one pixel.

We have demonstrated that the 4-D coherence of a non-stigmatic source such as illuminated blurred text can be sampled by the ACS. With the entire coherence sampled, we can digitally simulate the propagation of partially coherent light to apply an inverse filter and recover the unblurred image. We believe that these methods represent a powerful application of coherence theory to imaging and may ultimately benefit microscopy, astronomical imaging, and imaging through turbulence. Since detailed knowledge of the coherent transfer function of the system may not be required *apriori*, measurement of the 4-D coherence and subsequent decomposition may provide a powerful way of passively inferring details about intervening optical systems.

10

List of Figures

11

References

[1] O. Y. Kwon, "Real-time radial-shear interferometer," *Proc. SPIE*, vol. 551, pp. 32–35, 1985.

[2] R. N. Smartt and W. H. Steel, "Theory and application of point-diffraction interferometers (telescope testing)," *Jap. J. of Appl. Phys*, vol. 14, pp. 351–356, 1975.

[3] R. L. Kendrick, D. S. Acton, and A. L. Duncan, "Phase-diversity wave-front sensor for imaging systems," *Appl. Opt.*, vol. 33, no. 27, pp. 6533–6546, 1994.

[4] R. A. Gonsalves, "Nonisoplanatic imaging by phase diversity," *Opt. Lett.*, vol. 19, no. 7, pp. 495–497, 1994.

[5] E. Atad, J. W. Harris, C. M. Humphries, and V. C. Salter, "Lateral shearing interferometery. Evaluation and control of the optical performance of astronomical telescopes," *Proc. SPIE*, vol. 1236, no. 1, pp. 575–584, 1990.

[6] L. E. Schmutz, "Hartmann sensing at Adaptive Optics Associates," *Proc. SPIE*, vol. 779, pp. 13–17, 1987.

[7] R. Benedict, J. B. Breckinridge, and D. L. Fried, "Atmospheric compensation technology: Introduction," *J. Opt. Soc. Am. A*, vol. 11, no. 1, pp. 257–262, 1994.

12

[8] W. B. Bridges et al., "Coherent optical adaptive techniques," *Appl. Opt.*, vol. 13, no. 2, pp. 291–300, 1974.

[9] J. T. Salmon et al., "Adaptive optics package designed for astronomical use with a laser guide star tuned to an absorption line of atomic sodium," *Proc. SPIE*, vol. 2201, pp. 212–220, 1994.

[10] G. D. Love, "Wavefront control using a high-quality nematic liquid crystal spatial light modulator," *Proc. SPIE*, vol. 2566, pp. 43–47, 1995.

[11] D. Bonacinni, G. Brusa, S. Esposito, P. Salinari, and P. Stefanini, "Adaptive optics wavefront corrector using addressable liquid crystal retarders," *Proc. SPIE*, vol. 1543, 1990.

[12] A. P. Onokhov, V. V. Reznichenko, D. N. Yeskov, and V. I. Sidorov, "Optical wavefront corrector based on liquid crystal concept," *Proc. SPIE*, vol. 2201, pp. 1020–1026, 1994.

[13] J. Primot, G. Rousset, T. Marais, and J. C. Fontanella, "Deconvolution of turbulence-degraded images from wavefront sensing," *Proc. SPIE*, vol. 1130, pp. 29–32, 1989.

[14] V. Michau et al., "High-resolution astronomical observations using deconvolution from wavefront sensing," *Proc. SPIE*, vol. 1487, pp. 64–71, 1991.

[15] D. L. Marks, R. A. Stack, and D. J. Brady, "Astigmatic coherence sensor for digital imaging," *Opt. Lett.*, vol. 25, no. 23, pp. 1726–1728, 2000.

[16] E. Wolf, "New theory of partial coherence in the space-frequency domain. Part I: spectra and cross-spectra of steady-state sources," *J. Opt. Soc. Am.*, vol. 72, no. 3, pp. 343–351, 1982.

[17] L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics*. Cambridge: Cambridge University Press, 1995.

[18] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, Maryland: Johns Hopkins University Press, 1996.

Reference

Astigmatic
Coherence
Sensor

Target

Distorting
Plastic Plate

Cylindrical
Lenses

Sensor
Array
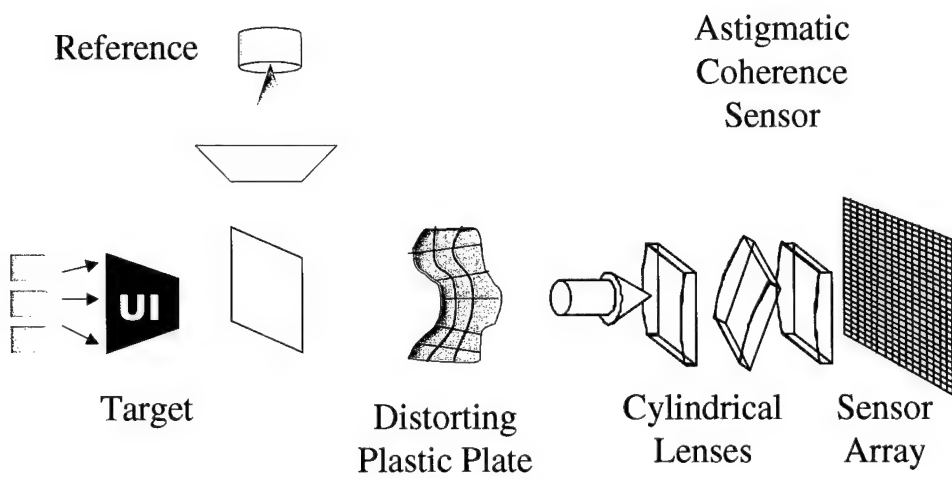
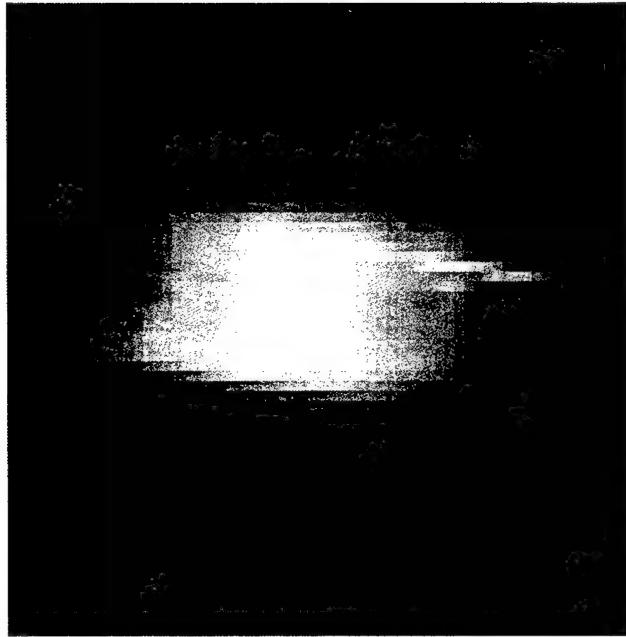Figure 1: Diagram of the setup of the source and Astigmatic Coherence
Sensor.

17

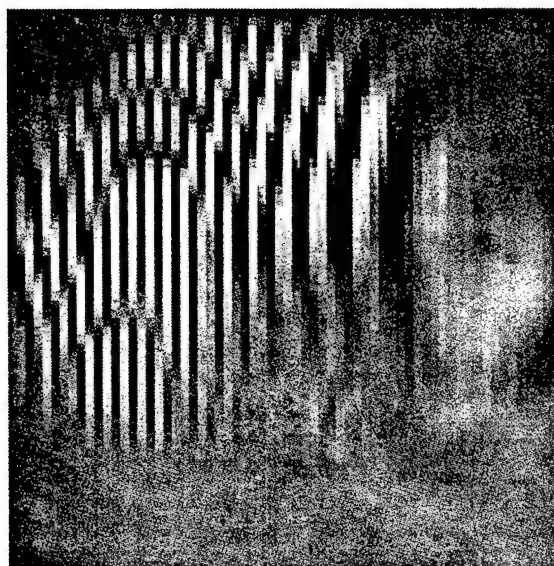Figure 2: Test object and reference viewed with a spherical lens with distortion.

19

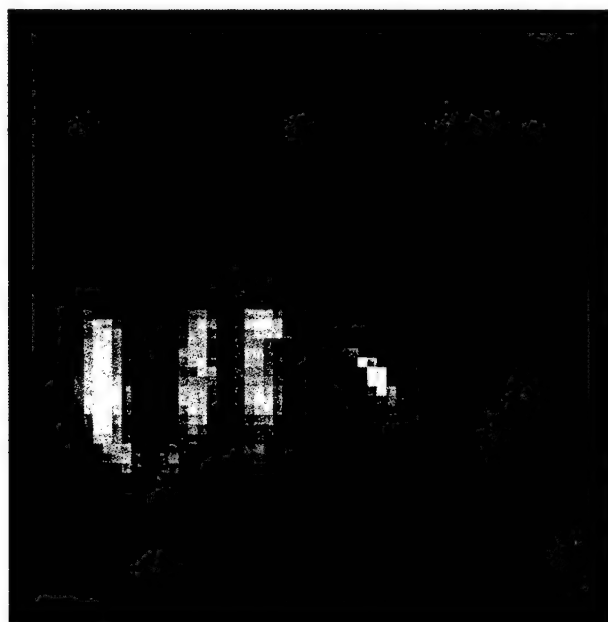Figure 3: Real part of the analytic field of the distortion as determined by coherent mode expansion.

Figure 4: Test object and reference viewed with a spherical lens with distortion.

Tomographic Analysis of Optical Images

Tomographic imaging of foam

M.R. Fetterman, E. Tan, L. Ying, R.A, Stack,
D.L. Marks, S. Feller, E. Cull, J.M. Sullivan,
D.C. Munson, Jr., S.T. Thoroddsen and D.J. Brady

Beckman Institute, University of Illinois at Urbana-Champaign,
Urbana, IL 61801

fetterma@uiuc.edu
http://www.phs.uiuc.edu

Abstract: The morphology of three-dimensional foams is of interest to physicists, engineers, and mathematicians. It is desired to image the 3-dimensional structure of the foam. Many different techniques have been used to image the foam, including magnetic resonance imaging, and short-focal length lenses. We use a camera and apply tomographic algorithms to accurately image a set of bubbles. We correct for the distortion of a curved plexiglas container using ray-tracing.

© 2000 Optical Society of America

OCIS codes: (100.6960) Tomography; (100.6950) Tomographic image processing

References and links

1. Denis Weaire, Stefan Hutzler, *The Physics of Foams*, (Oxford University, Oxford, 1999).
2. D. J. Durian, D. A. Weitz, D. J. Pine, "Multiple Light-Scattering Probes of Foam Structure and Dynamics," Science **252** 686 (1991).
3. C. Monnereau, M. Vignes-Adler, "Optical Tomography of Real Three-Dimensional Foams," Journal of Colloid and Interface Science **202** 45-53 (1998).
4. C. Monnereau, M. Vignes-Adler, "Dynamics of 3D Real Foam Coarsening," Phys. Rev. Lett. **80** (23) 5228-5231 (1998).
5. C. P. Gonatas, J. S. Leigh, A. G. Yodh, J. A. Glazier, B. Prause,"Magnetic Resonance Images of Coarsening Inside a Foam," Phys. Rev. Lett. **75** (3) 573-576 (1995).
6. H.P. Hiriyannaiah, "Computed Tomography for Medical Imaging," IEEE Signal Processing Magazine, 42-59, (March 1997).
7. L.A. Feldkamp, L.C. Davis, J.W. Kress, "Practical Cone-beam Algorithm," J. Opt. Soc. Am. A **1** (6) 612-619 (1984).
8. D. Marks, R.A. Stack, D.J. Brady, D.C. Munson Jr., "Visible Cone-Beam Tomography With a Lensless Interferometric Camera," Science **284** 2164-2166 (1999).
9. D.L. Marks, R.A. Stack, D.J. Brady, D.C. Munson Jr., "Cone-beam Tomography with a digital camera," Appl. Opt. (in review) 2000.
10. VTK Toolkit, http://www.kitware.com/vtk.html
11. H.K. Tuy, SIAM J. Appl. Math **43** 546 (1983).
12. M. Born, E. Wolf, *Principles of Optics*, (Cambridge University Press, Cambridge, 1980).
13. P. Soille, *Morphological Image Processing: Principles and Applications*, (Springer, Heidelberg, 1999).
14. S.A. Koehler, S. Hilgenfeldt, H.A. Stone, "A Generalized View of Foam Drainage: Experiment and Theory," Langmuir (http://pubs.acs.org/journals/langd5) **16** (15) 6327-6341 (2000).

1 Introduction

Light's interaction with soap bubbles creates colorful patterns that vividly illustrate the rudimentary principle of wave interference. But there is a lot more to learn about soap bubbles. In a cluster they serve as a model for many cellular systems occurring in nature. At low liquid content, they are organized into an intricate network of polyhedral foam adhering to certain geometric rules discovered by Plateau more than a century ago
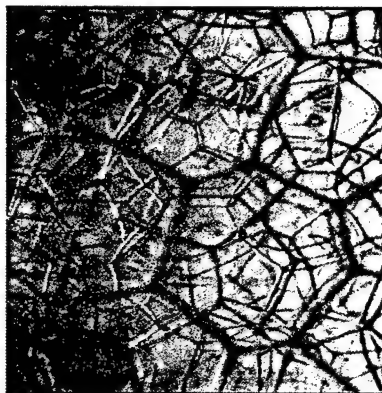
Fig. 1. A CCD video image of polyhedral aqueous foam showing the network of vertices and edges. The camera is set at a large depth of field to reveal the interior features.

[1]. Scientists are interested in how energy and entropy extremum principles determine the partition of space by soap bubbles. This motivates a technique capable of resolving the coordinates of vertices and edges of bubbles in foam.

Polyhedral aqueous foam made with soap solution looks like an open face structure because of soap film's transparency. The internal features are revealed to the extent that light rays can maintain straight paths before they are scattered and absorbed by the liquid borders. Fig. 1 shows the polyhedral network of vertices and edges captured by a video camera with a large depth of field.

Durian and his colleagues have developed a multiple light scattering technique to study foam [2]. It works by approximating light propagation through foam as a diffusion process. Light transmitted through a sample is measured and correlated with average bubble size. Durian's method has been demonstrated to work suitably with foam densely packed with small spherical bubbles of radius less than 50μm. However, it cannot provide information about the geometry of foam with polyhedral cells. To this end, attempts to measure the vertices' position and their connectivity would succeed by scanning a focal plane of a CCD camera, adjusted to small depth of field, through the layers of bubbles; internal features concealed in one direction are usually observable in other directions. Monnereau and Vignes-Adler have used this technique to reconstruct a cluster of up to 50 bubbles [3, 4]. The main disadvantage of this method is that image processing is needed to pick out the vertices from a set of noisy two-dimensional data slices. Vertices outside the focal plane are superimposed on data slices and thus make the resolution process difficult. Another approach to the foam imaging problem is to create a three-dimensional data volume by tomography. Magnetic resonance imaging (with tomographic algorithms) has been employed to examine the interior of foam with various degrees of success [5].

In this paper, we use Feldkamp's conebeam tomographic algorithms [6, 7, 8, 9] to reconstruct a three-dimensional foam. In the optical domain, we take many pictures of the object from different angles. These pictures are then processed by the conebeam algorithm to reconstruct the three-dimensional volume.

One advantage of this technique is that both the conebeam algorithm and the image processing both are designed to require no human intervention. This will speed up our rate of taking data, making it possible to analyze larger and more complex data sets. The time scale of this technique is also important. Currently we are taking 1 scan in a time of approximately 5 minutes, a time which could be reduced by taking images at video rate instead of still pictures. Since the time scale of the bubble development is on
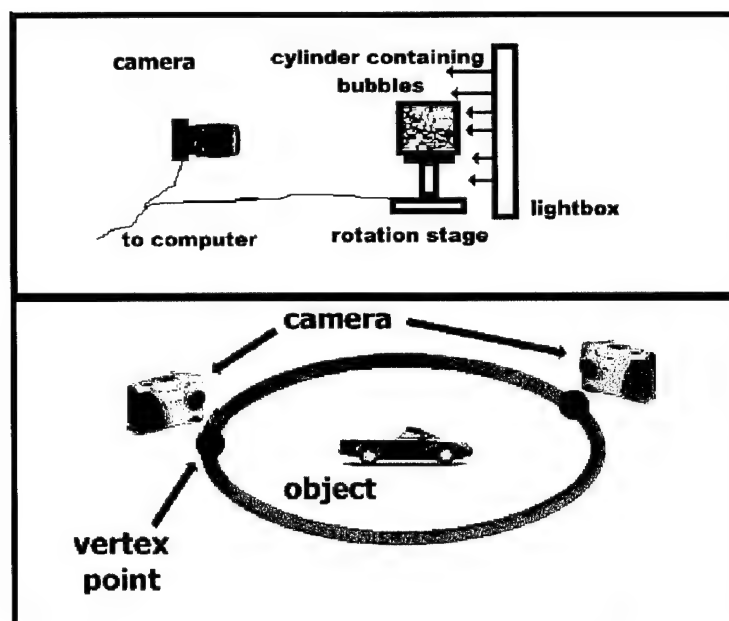
Fig. 2. *Top.* Experimental setup. The bubbles are in a cylindrical plexiglas container, that is placed on a rotation stage. The plexiglas container is held in a mount such that it is centered on the center of the rotation stage. A lightbox (a flat box with fluorescent lights that provides a uniform white light) is placed behind the cylinder, so that the camera sees the silhouetted image. The computer controls the rotation stage, and the computer also acquires images from the digital camera. *Bottom.* As shown above, the image rotates around on a stage, and the camera remains stationary. However, it is equivalent to view the object as stationary, while the camera rotates. The positions that the camera acquires an image at are referred to as the vertex points. The x, y, and z axes travel with the camera.

the order of hours, we will be able to take several images as the bubbles evolve.

An experimental problem that we encountered was that the container, a plexiglas cylinder, that held the bubbles distorted the ray path. Using ray-tracing we were able to compensate for this distortion. This distortion compensation may have applications to a wide range of tomography problems.

Our experimental results show a reconstruction of a test object, using the distortion correction algorithm. In future work, we will use a matched filter to extract the three-dimensional positions of the vertices.

2 Optical System Design and Algorithm

The experimental setup consists of an object mounted on a rotating stage. A schematic of the setup is shown in Fig. 2 *(top)*. The bubbles are in a cylindrical plexiglas container, that is placed on a rotation stage. The cylinder had an index of refraction of $n = 1.49$, an inner radius of $r_{inner} = 2.54cm$, and an outer radius of $r_{outer} = 3.17cm$. The plexiglas container is held in a mount such that it is centered on the center of the rotation stage. A lightbox (used in photography, the lightbox is a flat box with fluorescent lights that provides a diffuse white light) is placed behind the object, so that the camera records the silhouette of the object. The computer controls the rotation stage, and the computer also acquires images from the digital camera.

Our goal is to image the edges of the bubbles, which should show as lines on a silhouette image. Under optimal lighting conditions, the edges will appear as lines, while the faces will appear transparent. However, we do observe some scatter from the

Cone-beam tomography with a digital camera

Daniel L. Marks, Ronald Stack, Andrew J. Johnson, David J. Brady, and
David C. Munson, Jr.

We show that x-ray computer tomography algorithms can be applied with minimal alteration to the
three-dimensional reconstruction of visible sources. Diffraction and opacity affect visible systems more
severely than x-ray systems. For camera-based tomography, diffraction can be neglected for objects
within the depth of field. We show that, for convex objects, opacity has the effect of windowing the
angular observation range and thus blurring the reconstruction. For concave objects, opacity leads to
nonlinearity in the transformation from object to reconstruction and may cause multiple objects to map
to the same reconstruction. In x-ray tomography, the contribution of an object point to a line integral
is independent of the orientation of the line. In optical tomography, however, a Lambertian assumption
may be more realistic. We derive an expression for the blur function (the patch response) for a Lam-
bertian source. We present experimental results showing cone-beam reconstruction of an incoherently
illuminated opaque object. © 2001 Optical Society of America
   OCIS codes: 100.6950, 100.6890, 110.6880, 150.6910.

1. Background

Three-dimensional (3-D) imaging has important ap-
plications to machine vision, radiometry, animation,
modeling, microscopy, and source characterization.
3-D imaging has been implemented by machine-
vision techniques, such as stereo vision,[1] depth by
defocus,[2-4] and structured illumination and depth
cues,[1] and physical optics techniques, such as confo-
cal microscopy,[5] coherence tomography,[6,7] lidar,[8] and
coherence imaging.[9,10] Although each of these tech-
niques has been successful in certain situations, none
is universally applicable. We consider here a strat-
egy for reconstruction of incoherent sources based on
x-ray computer tomography. Although our tech-
nique is also not completely general, it requires only
weak assumptions about the nature of the source and
illumination when compared with typical machine-
vision systems, and it reconstructs over a wider solid
angle and with greater 3-D resolution when com-
pared with typical physical optical systems. The

drawbacks of our approach are that it involves poten-
tially unrealistic assumptions about opacity and
transmittance, that it works well only over the finite
depth of field of a planar imaging system, that it is
computationally intensive, and that it reconstructs
full volume data even for surface objects. Neverthe-
less, cone-beam tomography is a useful benchmark
for the potential of 3-D visible imaging.

Our goal is to estimate the 3-D radiant power den-
sity of an incoherent or Lambertian source. We
achieve this goal in two steps. First, we use a lens-
based camera to gather images of the source from
many perspectives. We interpret the images as lin-
ear projections through the source. Second, we ap-
ply a cone-beam tomography algorithm to these
projections to estimate the 3-D power density of the
source.

Computer tomography[11] is the reconstruction of a
source or scatterer from integrals along lines or
planes projected through the source volume. Stan-
dard approaches are grouped into parallel-beam, fan-
beam, and cone-beam techniques. Parallel-beam
systems measure the object transmittance for plane-
wave illumination. Fan-beam systems measure the
transmittance along rays projected from a point
source in a planar slice. The rays are confined to a
plane by a slit between the point source and the
object. Cone-beam systems measure the transmit-
tance along rays projected in a 3-D solid angle from a
point source. For a variety of technical and safety
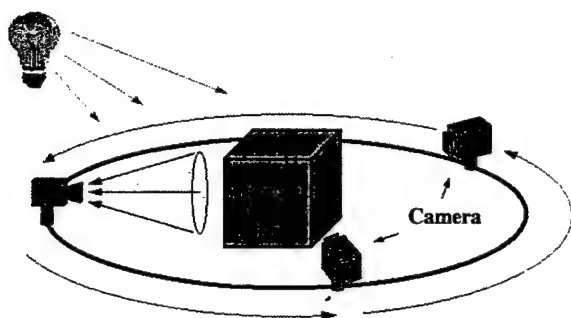reasons, x-ray systems until recently have been con-

Fig. 1. Optical imaging in cone-beam tomography.

fined primarily to fan-beam geometries. The constraints that promote fan-beam geometries for x rays do not apply in the visible, however, because visible illumination is relatively safe and large electronic visible sensor planes are ubiquitous. In short, visible systems are better suited to cone-beam algorithms.

In cone-beam tomography, linear projections through an object are gathered in discrete sets, where each set shares a common point of intersection, called the vertex point. A two-dimensional (2-D) set of projections is gathered at each vertex point, parameterized by the direction vector of rays projecting from the vertex point. To acquire 3-D data, the point moves along a prescribed path, called the vertex path, gathering projections that pass through each point along the path. Figure 1 shows the circular vertex path of a sensor around an object volume. A complete 3-D cone-beam data set is parameterized by position along the vertex path and projection direction. Many different algorithms have been developed to reconstruct a 3-D volume from cone-beam data. Exact reconstruction is possible for certain vertex paths, although incomplete paths are often used for implementation simplicity. In this paper we do not consider the relative merits of different algorithms. Rather, we show that x-ray cone-beam techniques can be applied with minimal modification to visible spectrum imaging.

The visible imaging problem differs from the x-ray problem in a number of respects. Whereas x-ray systems generally require artificial illumination, visible systems rely on ambient or spatially and spectrally incoherent sources. Whereas x-ray systems achieve acceptable resolution without compensating for diffraction, diffraction compensation by lenses and mirrors is readily available and highly desirable in the visible spectrum. Whereas x-ray targets are quasi-transparent, most visible targets are opaque. X-rays are difficult and dangerous to generate and detect; visible sources and detectors are well developed and safe. These differences lead to changes in the implementation of cone-beam imaging for the visible spectrum. In contrast with x-ray systems in which the vertex point coincides with a point source, in Sections 2 and 3 of this paper we describe how the vertex point for self-luminous or ambiently illumi-

nated incoherent visible imaging can be associated with a point in the principal plane of a visible lens and sensor assembly (e.g., a camera). In the x-ray system the source point and the sensor plane move in tandem around the vertex path on opposite sides of the object. In the visible system only the camera moves about the object. In Section 4 we consider the opacity issue and derive an expression for the patch response of a Lambertian source. In Section 5 we discuss criteria for selecting sampling points along the visible vertex path and consider the resolution one can expect in the reconstruction. In Section 6 we describe experimental demonstrations of the methods we consider.

Use of tomography to determine the shape and radiant intensity of a visible source may seem inefficient, especially given that many incoherent and all Lambertian sources are surface radiators. Ultimately a 2-D surface map may be all that is desired, yet cone-beam tomography requires a 3-D set of projections to determine this surface. In machine-vision applications economy is paramount, and this method may seem impractical. However, increasingly complex and disconnected surfaces begin to resemble volume objects. This method may be utilized on transparent and semitransparent visible sources such as biological tissues, so it has sufficient versatility and generality to be used with many instruments such as microscopes. Also, use of tomography avoids the need to register landmark points between the images; the only information needed is the position and orientation of the camera in space for each image. Tomographic algorithms are pure linear inversions and do not require decision making for computation. Because the end result is a 3-D power-density estimate, the algorithm makes no decisions about the location of surfaces. The patch response described below, however, can be combined with existing segmentation algorithms to abstract surfaces from volumetric data. Although perhaps data intensive and computationally expensive, we believe that the decreasing cost and increasing ubiquity of CCD and complementary metal-oxide semiconductor (CMOS) imagers, as well as faster computers, will increasingly justify robust, flexible, and general-purpose 3-D techniques. This could favor cone-beam tomography as a practical approach to image fusion.

2. Cameras and Cone-Beam Projections

The goal of cone-beam tomography is to reconstruct the 3-D source density function $D(\mathbf{r})$ from projections through a set of vertex points. The projections can be expressed as

$$ p(\mathbf{t}, \boldsymbol{\beta}) = \frac{1}{|\boldsymbol{\beta}|} \int D(\mathbf{t} + \alpha\boldsymbol{\beta}) d\alpha, \qquad (1) $$

where $\mathbf{t}$ is the vertex point and $\boldsymbol{\beta}$ is a vector along the projection direction. The geometry of these projections is shown in Fig. 2. A family of projections is captured at each vertex point. Several optical sys-
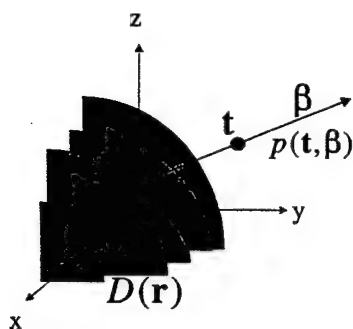
Fig. 2. Coordinate system that defines the projections. **t** is the vertex point for a family of projections and **β** is a direction vector along a particular projection.

tems capture projections of this approximate form. In previous research, we considered tomographic reconstruction from projections of this type using pinhole cameras, coherence imagers, and cubic-phase-aberrated systems.[9,12,13] Although any of these systems might be used with cone-beam algorithms, in this paper we focus more on the applicability of the algorithms themselves to visible imaging and less on optical design issues. Accordingly, we limit our attention here to imaging with a standard lens-based camera.

Consider the camera geometry shown in Fig. 3. An object a distance z from the input principal plane of the camera is imaged a distance z' behind the output principal plane. Newton's equations[14] relate the ratio of the image cross section y' to the object cross section y as $y'/y = f/(z + f) = (z' + f')/f'$. f and f' are the focal lengths of the source and image spaces, respectively. From these equations one concludes that $y'/z' = -(f/f')(y/z)$ or that the slope of the angle the image subtends to the optical axis is proportional to the slope the source subtends to the axis. The images of sources that share the same slope y/z will have their images centered at the same position on the image plane. Each sensor location on the image plane can be considered to measure the total power from all the sources along the same slope y/z, which is a projection of the source that intersects the center of the source principal plane. The projections then form a cone, so that the image plane measures the cone-beam projections of the source, with the vertex point being the center of the source principal plane. Under the assumption that all source
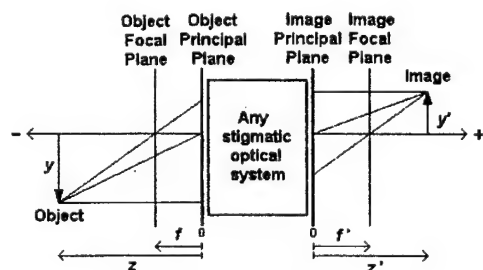
points form point images, the image on the output plane of the camera is

$$p(0, \mathbf{r}) = \int_0^\infty D(\kappa z \rho_x, \kappa z \rho_y, z) \mathrm{d}z, \tag{2}$$

where ρ is the position on the image plane and $\kappa = -f'/fz'$. Equation (2) corresponds to Eq. (1) for the vertex point $\mathbf{t} = 0$ and $\beta = (\kappa \rho_x, \kappa \rho_y, 1)$. We obtain a complete set of vertex points and projections by moving the camera around the object volume. We note that these projections are already weighted, so projections specified this way do not need the weighting factor of Eq. (7).

Our derivation of Eq. (2) neglects the fact that Newton's equations are not satisfied for a fixed image plane as we vary the object range. When the object and image planes satisfy the Newton equations, all rays from a point on the object converge to a single point on the image. When we shift the object plane without shifting the image plane, rays from an object point no longer cross in the image plane, and the image becomes defocused. The longitudinal range over which the object planes are approximately in focus for a given image plane is termed the depth of field. Regions of the source that are outside the depth of field will be so blurred that Eq. (2) cannot be considered accurate. We can address this discrepancy by adjusting the camera aperture size or apodization to extend the depth of field or by using high depth-of-field imagers, such as the cubic-phase aberration[15,16] or interferometric imagers.[9]

To consider approximate depth of field in more detail, we rederive Eq. (2) using diffraction integrals. For simplicity we limit our attention to a thin-lens imaging system. For such a system, the optical image field $\psi(\mathbf{r}')$ at a distance z' behind the lens that is due to the object field $\psi_D(\mathbf{r})$ a distance z in front of the lens is

$$\psi(\mathbf{r}') = \exp\left(j\frac{\pi|\mathbf{r}'|^2}{\lambda z'}\right) \int \exp\left(j\frac{\pi|\mathbf{r}|^2}{\lambda z'}\right)$$
$$\times \psi_D(\mathbf{r})h\left(\mathbf{r} + \frac{z}{z'}\mathbf{r}', z\right)\mathrm{d}\mathbf{r}, \tag{3}$$

where

$$h(\mathbf{r}, z) = \int t(\mathbf{r}'')\exp\left(j\frac{2\pi}{\lambda z}\mathbf{r}\cdot\mathbf{r}''\right)$$
$$\times \exp\left[j\frac{\pi}{\lambda}\left(\frac{1}{z} + \frac{1}{z'} - \frac{1}{f}\right)|\mathbf{r}''|^2\right]\mathrm{d}\mathbf{r}''. \tag{4}$$

$t(\mathbf{r}'')$ is the pupil function for the lens aperture and f is the focal length of the lens. With no aberration, $t(\mathbf{r}'') = 1$ inside the aperture and zero outside the aperture. The camera detects the intensity on the image plane $P(\mathbf{r}') = \langle|\psi(\mathbf{r}')|^2\rangle$. If the object field is

Fig. 3. General stigmatic optical system.

incoherent such that $\langle\psi_D(\mathbf{r}_1, z_1)\psi_D{}^*(\mathbf{r}_2, z_2)\rangle = D(\mathbf{r}_1)\delta(\mathbf{r}_1 - \mathbf{r}_2)$, then the detected signal is

$$P(\mathbf{r}') = \int D(\mathbf{r}) \left| h\left(\mathbf{r} + \frac{z}{z'}\mathbf{r}', z\right) \right|^2 d\mathbf{r}, \qquad (5)$$

where $d\mathbf{r}$ represents a 3-D integral over the transverse components in $\mathbf{r}$ and the longitudinal component z.

In going from Eq. (3) to Eq. (5) we added an integral over z on the assumption that $D(\mathbf{r})$ is the primary source radiance or scattering efficiency of the object, which is not generally equal to the optical intensity at $\mathbf{r}$. In summing the contributions to $P(\mathbf{r}')$ over z we assume that the camera response is linear in intensity. Equation (5) is a correct but not necessarily a unique means of expressing the focal plane projections in terms of the source density. Equation (5) generally cannot be inverted to recover $D(\mathbf{r})$. To see this, one need only note that any allowed $P(\mathbf{r}')$ can be obtained simply by making the intensity in the in-focus plane proportional to the target focal plane value. If, however, Eq. (5) can be viewed as a projection through the vertex point $\mathbf{r}' = 0$, then it can be inverted with projections through other vertex points to unambiguously recover $D(\mathbf{r})$. In principle, the measurement of $P(\mathbf{r}')$ along a vertex path can determine $D(\mathbf{r})$ even if it cannot be reduced to cone-beam projections, but we do not consider this possibility in this paper.

Equation (5) is equivalent to Eq. (2) if $h(\mathbf{r}, z) \cong \delta(\mathbf{r})$. In this case the assumption that the camera captures cone-beam projections is valid. Referring to Eq. (4), we can see that $h(\mathbf{r}, z)$ is an approximate delta function so long as

$$\frac{A^2}{\lambda}\left(\frac{1}{z} + \frac{1}{z'} - \frac{1}{f}\right) < 1,$$

where A is the aperture of the lens. This implies a depth of field in the object space of $\Delta z \approx F^2\lambda$, where $F = z/A$ is the f-number for the imaging system. One can always improve the depth of field by reducing the aperture, but this course also reduces the transverse resolution by blurring $h(\mathbf{r}, z)$. In the remainder of this paper, we assume that the object space lies completely within the depth of field and that Eq. (2) is valid.

We assume that our camera captures cone-beam projections covering the object space as it moves along a vertex path. The vertex point $\mathbf{t}$ for a given camera position is the center of the principal plane of the camera's imaging system. Projections are captured on the camera's focal plane for rays along $\boldsymbol{\beta} = (\rho_x'/z', \rho_y'/z', 1)$. This approach to gathering projections differs from the standard approach for x-ray cone-beam tomography. In x-ray systems, the vertex point corresponds to the position of a point source, and projection data are measured by a sensor plane on the opposite side of the object from the point source. In the visible systems described here, the object is either ambiently illuminated or self-

luminous, and the vertex point and the sensor plane are on the same side of the object. The basic geometry for our system is illustrated in Fig. 1.

3. Cone-Beam Image Formation

Having decided that a camera gathers cone-beam data, we must consider how to invert this data to reconstruct the source density. A number of cone-beam algorithms have been developed over the past two decades, and algorithm development continues to be an active area of research. Algorithms can be classified according to the nature of the vertex path they accept and the approach to data inversion. Tuy[17] showed that exact inversion is possible for vertex paths such that all planes slicing the object volume also intersect the vertex path. Even for paths that technically satisfy Tuy's condition,[17] inversion fidelity is limited in practice because both the vertex path and the projection space are sampled discretely. In many cases, even vertex paths that do not satisfy Tuy's condition provide satisfactory results. The most common inversion algorithms use convolution backprojection methods, but a number of alternatives to this approach exist. Continuing developments are improving the computational efficiency, vertex path tolerance, and sampling efficiency of cone-beam systems.

Our goal in this paper is simply to show how cone-beam inversion applies to visible imaging systems. We do not compare or analyze potential inversion algorithms. Accordingly, we chose to demonstrate imaging using the simplest and most popular algorithm, as developed by Feldkamp et al.[18] Feldkamp's algorithm is a filtered backprojection algorithm and can be considered an extension of the 2-D fan-beam reconstruction algorithm to 3-D cone-beam reconstruction. The algorithm uses the circular vertex path illustrated in Fig. 1. As the sensor is moved in a circle around the object space origin, it rotates to point toward the origin. The plane of the vertex path is called the midplane. Feldkamp also considers the plane of rotation, which is the plane orthogonal to the midplane and to the ray from the vertex point to the origin containing the origin and the axis of rotation. The geometry of the midplane and plane of rotation are illustrated in Fig. 4. Feldkamp parameterizes each projection by the vertex point and by its point of intersection with the plane of rotation. As illustrated in Fig. 4, the vertex point is parameterized by the angular position on the vertex path Φ. The intersection with the plane of rotation is also shown. Feldkamp parameterizes this point with the variables Y and Z. Because a camera measures the projections parameterized by the angle subtended to the optical axis, we replace Y and Z with $\xi_y = Y/d$ and $\xi_z = Z/d$. Other than this change, our notation is identical to Feldkamp's formulation. In Feldkamp's algorithm, the angles of the projections that the sensor gathers are spaced such that the horizontal and vertical slopes of the projections, relative to the sensor's rotational angle and position, are sampled at regular intervals. A paraxial imaging sys-
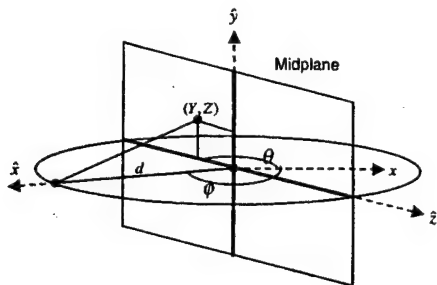
Fig. 4. Illustration of coordinate systems used on midplane in Feldkamp's algorithm. $\hat{x}$ and $\hat{z}$ are the dimensions in the midplane, with $\hat{x}$ pointing at the vertex point and $\hat{z}$ pointing along the midplane; $\hat{y}$ points out of the midplane; Y and Z are the coordinates on the plane of rotation; d is the radius of the vertex circle; ϕ is the angle from the x axis to the vector from the origin to the vertex point; and θ is the angle between the x axis and the $\hat{z}$ axis.

tem has the same regular spacing, assuming the image plane sensor is a rectilinear grid of pixels.

Inversion by Feldkamp's algorithm consists of two steps. First, the projection data $P_\phi(\xi_y, \xi_z)$ are weighted and convolved with the separable filter functions

$$g_y(\xi_y) = \int_{-\omega_{y0}}^{\omega_{y0}} d\omega |\omega| \exp(i\omega\xi_y - 2|\omega|/\omega_{y0}),$$

$$g_z(\xi_z) = \frac{\sin(\xi_z\omega_{z0})}{\pi\xi_z}. \qquad (6)$$

The bandwidth parameters ω_{y0} and ω_{z0} are given by the sampling period on the camera. This produces the intermediate function

$$\bar{P}_\phi(\xi_y, \xi_z) = \int_{-\infty}^{\infty} d\xi_y' \int_{-\infty}^{\infty} d\xi_z' g_y(\xi_y - \xi_y') g_z(\xi_z - \xi_z')$$

$$\times P_\phi(\xi_y', \xi_z')(1 + \xi_y'^2 + \xi_z'^2)^{-1/2}. \qquad (7)$$

Finally, the source density is estimated as the backprojection of the filtered projection data according to

$$D_E(\mathbf{r})$$

$$= \frac{1}{4\pi^2} \int \frac{d^2}{(d + \mathbf{r}\cdot\hat{x}')^2} \bar{P}_\phi\left[\frac{\mathbf{r}\cdot\hat{y}'}{d + \mathbf{r}\cdot\hat{x}'}, \frac{\mathbf{r}\cdot\hat{z}'}{d + \mathbf{r}\cdot\hat{x}'}\right] d\phi, \qquad (8)$$

where d is the radius of the vertex path and the unit vectors of the coordinates that rotate with the camera are $\hat{x}'$, $\hat{y}'$, and $\hat{z}'$. The reconstructed power density $D_E(\mathbf{r})$ is a function of the position $\mathbf{r}$ in the source space, and it is an estimate of the original power density $D(\mathbf{r})$. We modified the filtering function $g_y(\xi_y)$ from the original Feldkamp algorithm by windowing it in the frequency domain by an exponential function to prevent ringing at the edges.

4. Opacity and the Patch Response

So far we have assumed that the source is transparent and incoherent. In this case, a lens-based cam-

era will measure cone-beam linear projections of the source, from which we can infer the power density of the source using cone-beam tomographic algorithms. However, most sources at optical frequencies do not satisfy the transparent and incoherent assumption. Often, more realistic sources are opaque, some are reflecting rather than self-luminous, and many such sources have rough surfaces. A large class of realistic optical sources consists of Lambertian surface radiators. This class includes diffuse light sources and rough surface objects illuminated by light having low spatial coherence. Unlike incoherent sources, points on the surfaces of Lambertian sources do not radiate isotropically. Because of the anisotropy and opacity of Lambertian sources, lens-based imagers do not measure linear projections of such sources. If one samples the wave front from a Lambertian object with a lens-based camera and uses these samples to reconstruct the object using a cone-beam algorithm, the resulting power-density reconstruction is flawed but is related in a predictable way to the actual surface radiant intensity of the source. Although the power-density reconstruction is not perfect, many features of the source are preserved and can be found by filtering the reconstruction. By studying the effect of Lambertian anisotropy on the reconstructed power density, we can account for these effects in the power-density reconstruction.

A Lambertian radiating source can be regarded as consisting of infinitesimal surface patches, each of which radiates according to Lambert's law, which states that the radiated power at the angle θ from the patch surface normal is proportional to $\cos \theta$. If the Lambertian source is also convex, then no surface patches are occluded from any angle, so that each patch is visible from a hemisphere of directions. Under these restrictions, a lens-based camera will detect an incident intensity that is a linear function of the surface radiant power density of the source. The measured intensity from a given patch will depend only on the relative positions and angular orientations of the patch and camera and will not depend on occlusion from any parts of the source. Each surface patch then provides an individual contribution to the cone-beam reconstruction that depends only on its position and orientation. We define the contribution to the reconstruction from a given infinitesimal patch as the patch response, which is the 3-D power-density reconstruction of a patch as the patch size approaches zero. This patch response is a characteristic of the Lambertian radiation pattern of sources, and it is a function of both the position and the orientation of the patch. To find the reconstructed power density of a convex Lambertian source, one needs to integrate the patch responses of all the surface patches, with each patch response appropriately translated and rotated to match the position and orientation of each patch and weighted according to the power density of each patch.

The patch response is a characteristic blur function attached to each patch of the reconstructed source. To investigate this blur, we calculate the patch re-
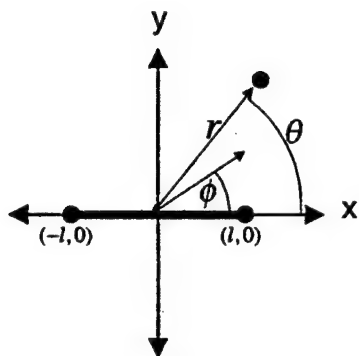
Fig. 5. Coordinate system used for the projection-slice theorem 2-D integral. $2l$ is the length of the Lambertian line, r and θ are the polar coordinates of the 2-D density reconstruction, and ϕ is the angle of the projections of the Lambertian line.

Fig. 6. Density plot of a 2-D patch response function.

sponse for a 2-D Lambertian source. We consider a unit radiant intensity thin-line Lambertian source of length $2l$ centered at the origin and aligned horizontally. The reconstructed power density of the Lambertian line source will be that of a finite extent line patch, from which we can determine the patch response by taking the limit of the power density, normalized by the length $2l$, as the patch size approaches zero. Figure 5 shows the line and the coordinates of the projections and the patch response. The Fourier projection-slice theorem can be used to determine the power density of a 2-D source from its parallel-beam projections. This derivation is detailed in Appendix A. The derivation consists of one computing the response of a Lambertian line of length $2l$ using the Fourier projection-slice theorem and then taking the limit of this response as the line length approaches zero. The resulting patch response is

$$\mathrm{PRF}(x, y) = \frac{1}{4\pi^2} \left\{ \frac{2x}{(x^2 + y^2)^{3/2}} \log\left[\frac{(x^2 + y^2)^{1/2} + x}{(x^2 + y^2)^{1/2} - x}\right] - \frac{4}{x^2 + y^2} \right\}, \tag{9}$$

where PRF is the patch response function and x, y are coordinates transverse to the patch surface normal. This function is plotted in Fig. 6. One may be tempted to filter (correlate) the reconstruction with this patch response to partially remove its effect on surfaces. However, because the patch response is oriented according to the surface normal at each surface point, it is actually a class of functions that is parameterized by a rotation angle. To deconvolve the patch response, the surface orientation must be known or estimated. Fortunately, the patch response still has a local, confined nature so that the reconstructed power density does not stray too far from the original surface.

For nonconvex objects, additional complications arise because parts of the object obscure other parts. Some patches will be visible from a restricted (and possibly disjointed) set of angles, less than the full 180 deg. The power-density reconstruction of a
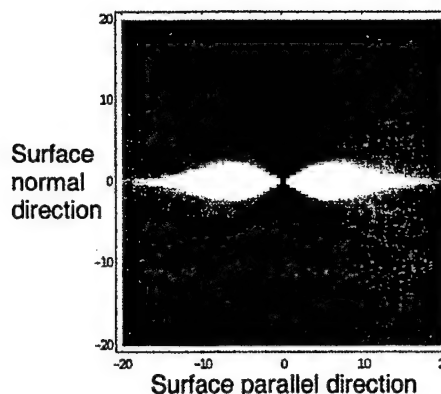
patch will then become dependent on the set of angles it is obscured from. Tomographic reconstruction from a limited range of angles is a well-studied problem. An obscured patch will blur normal to the directions it is obscured from. Similarly, if projections of the source are gathered from a restricted range of angles, the patch response will become orientation dependent, and similar blurring will occur.

The reconstruction of arbitrarily shaped Lambertian sources that have a constant surface brightness is similar to the reconstruction of an object from its silhouettes. The intensity a lens-based camera would measure of a constant-brightness Lambertian source in front of a uniform dark background is just the inverted silhouette, with the bright area indicating the shadow of an opaque object having the same shape as the source. If there exists a convex Lambertian source with the same silhouettes as a nonconvex source, then the power-density reconstruction of the nonconvex source will be identical to that of the convex source because they are indistinguishable from their projections. Figure 7 shows nonconvex and convex Lambertian sources that have the same projections and therefore the same power-density reconstruction. Only the extreme points on the object that bound the silhouettes are determinable from the projections, as shown in Fig. 7, and these points are common to all the sources that share the same silhouettes. The patch response power density has the property of concentrating power density around these points, because the patch response tends to cancel itself on straight edges and accentuate itself at cor-
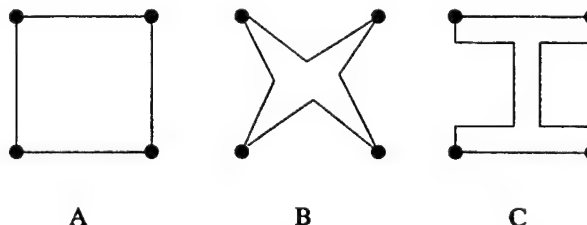

A   B   C

Fig. 7. Three constant-brightness Lambertian objects that have the same silhouettes and therefore the same reconstructed power density.
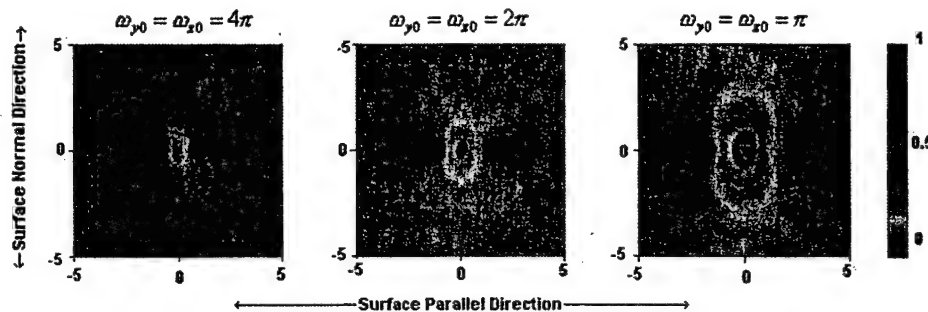
Fig. 8. Density plot of the patch response with various sampling densities.

ners when convolved with a convex surface. However, if the source has a nonconstant brightness, the power density will resolve the nonedge points.

So far, we have considered only 2-D Lambertian sources in the context of the patch theory. 3-D sources introduce an additional complication. In 3-D space, a four-dimensional set of linear projections can be measured (parameterized, for example, by position on a surface and direction), but normally only a 3-D subset of these would be necessary for tomographic reconstruction of a nonscattering source in a bounded volume. In the case of a nonscattering source, a sensor at any vertex point measures projections from the entire source because occlusion does not occur, so the set of projections needed to reconstruct the volume is dependent only on the shape of the support volume and not its contents. 3-D Lambertian sources can have occlusion, so the vertex path of a cone-beam algorithm must be chosen to sample sufficiently projections originating from all surfaces of interest in the source. Because the subset of the projections sampled from any given patch on the source is determined by the vertex path of the sensor, the patch response of a given patch will depend on the vertex path. Also the reconstruction can depend on the cone-beam tomographic algorithm used because of differences in how the algorithms reconstruct incomplete data or handle the space-variant point-spread function of the reconstruction. Rather than describe the general 3-D patch response that accounts for the vertex path, we confined the vertex path for our experimental setup to a plane containing the source, but sufficiently far away from the source, so that the 2-D patch response can be used in approximation.

As suggested above, the power-density reconstruction of a Lambertian source can be filtered to enhance features of the source that are blurred because of the patch response. Although the patch response for any given patch depends on its orientation, so that a matched filter should match this same orientation, a radially symmetric filter could match all orientations imperfectly and serve as a useful compromise. We chose to use a Laplacian filter because the $(x^2 + y^2)^{-2}$ dependence of the patch response suggests that a Laplacian filter might work well to transform the patch response into a deltalike function, as it does for $(x^2 + y^2)^{-1}$.

5. Sampling, Aperture, and Resolution

The Feldkamp cone-beam reconstruction method and the patch response derived above are for continuous sets of projections and vertex paths. In real data acquisition, one gathers discretely sampled sets of projections from a finite number of points on the vertex path. Although we do not rigorously consider the effects of sampling, we present examples of how finite sampling affects the patch response, and we give a heuristic argument that provides an estimate of the sampling requirements needed to produce a satisfactory reconstruction.

In optical cone-beam tomography, the aperture of the camera and the finite number of pixels on the camera determine the resolution of the image recovered. Each pixel corresponds to a cone-beam projection of the source, so a denser pixel array, with a correspondingly larger aperture, results in a higher density of sampled projections. A more densely sampled set of projections will improve the image quality by increasing the spatial bandwidth of the reconstructed patch, resulting in an improved patch response. Figure 8 shows an example of the effect of our sampling at various rates, where ω_{y0} and ω_{z0} are the spatial sampling rates in the y and z directions, respectively. To achieve one-pixel resolution, one must sample at a resolution of one pixel per projection at the midplane, which is achieved by the sampling rate $\omega_{y0} = \omega_{z0} = 2\pi$. Any higher rate unnecessarily oversamples the patch response, and lower rates do not sufficiently sample the patch response to achieve one-pixel resolution. Conversely, the sampling rate of the camera sets the achievable bandwidth and resolution of the reconstruction. In our experiments, we set the resolution and field of view of the cameras to maximize the resolution while maintaining the requirement of fitting the source within the field of view at all vertex points.

A second consideration is the number of vertex points required to produce adequate resolution everywhere in the reconstruction zone. For Feldkamp's algorithm, vertex points are equally spaced on a circular path around the reconstruction volume. For a 128×128 pixel reconstruction on the plane of the vertex circle, Fig. 9 shows the quality of the patch response by use of various numbers of projection angles. As the number of projections approaches the
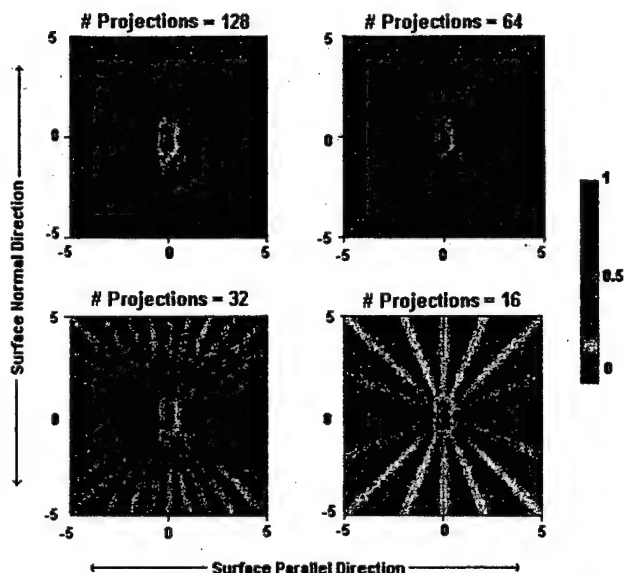
Fig. 9. 2-D patch response with various projection angles.

number of pixels in the midplane, the radial streak artifacts that are due to the sampling of the vertex path disappear. Denser sampling produced only slight improvement in reconstruction quality, which agrees with Ref. 19 for the 2-D parallel beam case. We set the number of projections equal to the number of pixels across the midplane, or 128 for our 128 × 128 reconstruction.

The following examples provide two simple rules that can be used to avoid the artifacts shown here. Select the imaging resolution on the camera as one projection per pixel at the midplane, and set the vertex path sampling density with the number of projections equal to the number of pixels across the midplane.

6. Experimental Verification

We tested the idea of reconstructing a Lambertian source using a cone-beam tomographic algorithm. The experiment consisted of rotating an object, taking an image from each angle, and applying Feldkamp's cone-beam algorithm to estimate the power density of the source. The object was a toy bear and
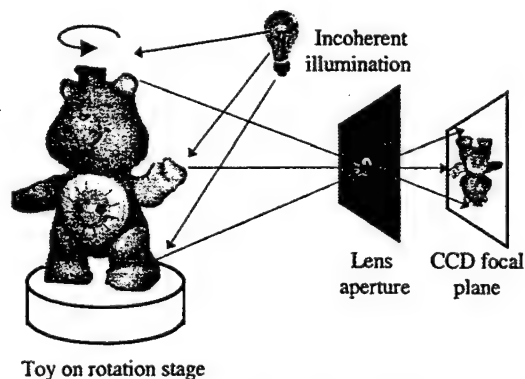

Fig. 10. Setup of cone-beam data acquisition.

was illuminated by a long white-light fluorescent tube lamp that was approximately 10 cm from the toy but not directly seen by the camera. Figure 10 illustrates the setup. The object was placed on a rotation stage in front of a black absorbing background, which was sequentially rotated through 360 deg in 128 equally spaced steps. At each position, a camera-lens system and computer recorded an image of the object, which was approximately 1 m away from the camera. The camera-lens system consisted of a 50-mm focal-length lens and a backilluminated 16-bit CCD focal plane array. To prevent nonlinearities in the data, the focal plane used no automatic gain control or saturation. The angular magnification and location of the principal plane of the lens were calibrated manually. Each frame taken by the camera was 512 × 512 in size. Only 256 × 256 pixels of the field of view were needed, and the data were decimated to 128 × 128. The data set, which included the 128 frames of the object taken from different angles, the angular magnification of the object, and the distance from the principal plane to the axis of the rotation stage, was processed as parameters of Feldkamp's cone-beam algorithm.

The results of our reconstruction are shown in Fig. 11. Figures 11(a) and 11(b) show a ray-cast rendering of the 3-D power-density reconstruction of the toy bear from two angles. Because the object can be shown from angles and positions where the object was not imaged originally, more information about the object's true shape can be visualized. Figure 11(c) shows a cross section of the reconstructed power density of the head of the bear. Figure 11 demonstrates that the surfaces are not sharply reconstructed; some power is effectively smeared inside the body. This imperfect reconstruction is a result of the patch response function contributing power into the interior of the reconstruction of the bear's head. The power density tends to be concentrated near the position of the original surface, so the surface can be still be located within the data set.

To reduce the power reconstructed inside the opaque object, we applied the Laplacian filter described at the end of Section 4. The new filtered power density is shown in Fig. 12. The features on the outer surface of the bear are now clearer because much of the power on the inside of the object has been removed. The cross section of the bear's head is shown in Fig. 12(c) after filtering. Compared with Fig. 11(c), the cross section is basically hollow except for the surface. This supports the idea that the filtered power density will more accurately represent the true surface. In ongoing research, we are studying more sophisticated methods for removing the effect of the nonideal patch response.[20]

Appendix A

In this appendix we derive the patch response given in Eq. (9). The goal is to derive the 2-D patch response. To do this, first we derive a general formula for all 2-D Lambertian objects. We insert the projections for a finite size Lambertian line object into
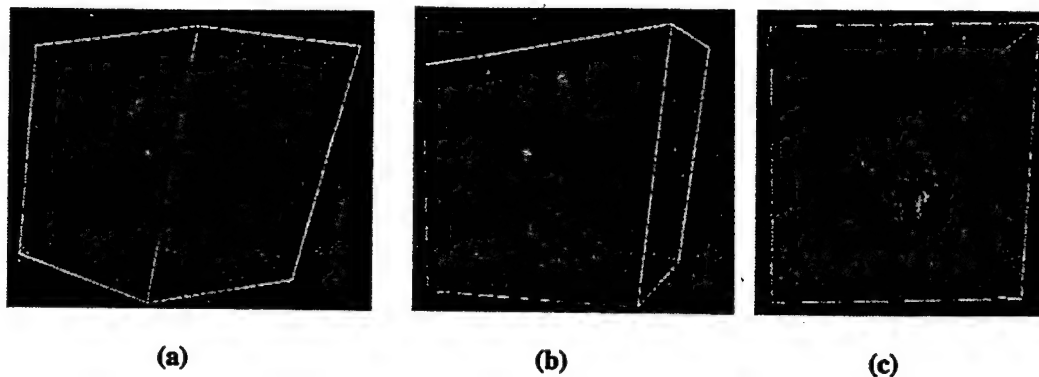
(a)   **(b)**   **(c)**

Fig. 11. Toy bear reconstructed by Feldkamp's algorithm: (a) and (b) ray-cast renderings of the side and front views of the power density of the bear, respectively; (c) a lateral cross section through the bear's head.

this formula and take the limit as the line length approaches zero. Finally, we integrate the projection angle to calculate the response to a infinitesimal patch. To begin, we define a 2-D Lambertian object with a set of extents $l(\phi)$. The extent $l(\phi)$ is defined by the maximum value of $r \cos(\theta - \phi)$ on the object for a given angle ϕ. For a Lambertian line source with extents $l(\phi)$, we define the projections $p(l, \phi) = 1$ for $-l(\phi + \pi) \leq l \leq l(\phi)$ and zero otherwise. Starting from the projection-slice theorem,

$$f(r, \theta) = \int_0^\infty \int_0^{2\pi} s \, \exp[i2\pi rs \cos(\theta - \phi) - ks]$$
$$\times P(s, \phi)\mathrm{d}\phi\mathrm{d}s, \qquad (A1)$$

where $P(s, \phi) = \int_{-\infty}^{\infty} p(l, \theta)\exp(2\pi sl)\mathrm{d}l$. We added an extra parameter $k \geq 0$ that keeps the reconstruction finite by bandlimiting the Fourier transform. For the Lambertian source with extents $l(\phi)$, $P(s, \phi)$ will be the Fourier transform of the projections

$$p(l, \theta) = \mathrm{rect}\{[l - l(\phi)/2]/l(\phi)]\}$$
$$+ \mathrm{rect}\{[l + l(\phi + \pi)/2]/l(\phi + \pi)\},$$

which is

$$P(s, \phi) = \frac{\sin \pi sl(\phi)}{\pi s} \exp[-i\pi sl(\phi)] + \frac{\sin \pi sl(\phi + \pi)}{\pi s}$$
$$\times \exp[i\pi sl(\phi + \pi)]. \qquad (A2)$$

We insert this into Eq. (A1) to obtain

$$f(r, \theta) = \int_0^\infty \int_0^{2\pi} s \, \exp[i2\pi rs \cos(\theta - \phi) - ks]$$
$$\times \left\{ \frac{\sin \pi sl(\phi)}{\pi s} \exp[-i\pi sl(\phi)] \right.$$
$$\left. + \frac{\sin \pi sl(\phi + \pi)}{\pi s} \exp[i\pi sl(\phi + \pi)] \right\}\mathrm{d}\phi\mathrm{d}s. \qquad (A3)$$

We can simplify this to obtain

$$f(r, \theta) = \int_0^\infty \int_{-\pi}^{\pi} \frac{\exp[i2\pi rs \cos(\theta - \phi) - ks]}{i\pi}$$
$$\times \{1 - \exp[i2\pi sl(\phi)]\}\mathrm{d}\phi\mathrm{d}s. \qquad (A4)$$

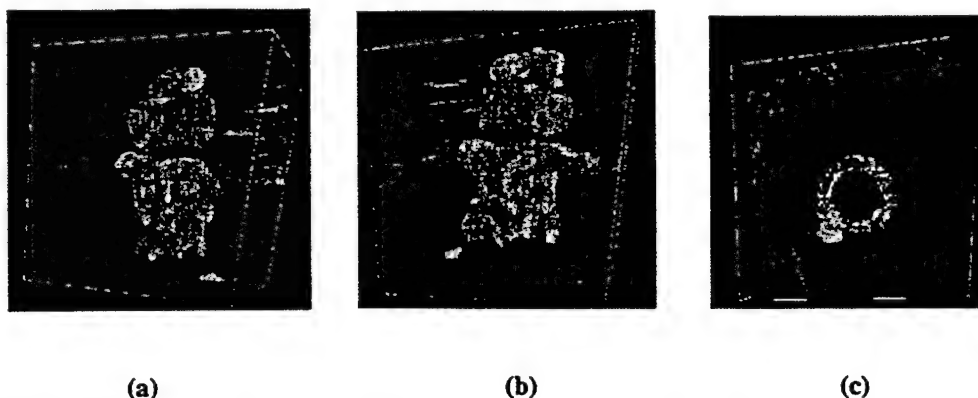

(a)   **(b)**   **(c)**

Fig. 12. Laplacian-filtered reconstructed bear: (a) and (b) ray-cast renderings of the side and front views of the power density of the bear, respectively; (c) a lateral cross section through the bear's head.

$$f(x, y) = \left(4[k^2 + 4\pi^2(x^2 + y^2)]^{1/2} + 2\pi x \log \frac{\{[k^2 + 4\pi^2(x^2 + y^2)]^{1/2} - 2\pi(x - y)\}^2 + k^2}{\{[k^2 + 4\pi^2(x^2 + y^2)]^{1/2} + 2\pi(x - y)\}^2 + k^2}\right.$$
$$\left. + 2\pi x \log \frac{\{[k^2 + 4\pi^2(x^2 + y^2)]^{1/2} + 2\pi(x + y)\}^2 + k^2}{\{[k^2 + 4\pi^2(x^2 + y^2)]^{1/2} - 2\pi(x + y)\}^2 + k^2}\right)[k^2 + 4\pi^2(x^2 + y^2)]^{-3/2}. \tag{A10}$$

The integration is performed to become

$$f(r, \theta) = \int_{-\pi}^{\pi} [2\pi^2 r \cos(\theta - \phi) + i\pi k]^{-1}$$
$$- \{2\pi^2[r \cos(\theta - \phi) - l(\phi)] + i\pi k\}^{-1} d\phi. \tag{A5}$$

The left-hand term integrates to zero, so we can eliminate it to obtain

$$f(r, \theta) = \int_{-\pi}^{\pi} \{2\pi^2[l(\theta) - r \cos(\theta - \phi)] - i\pi k\}^{-1} d\phi, \tag{A6}$$

which is the general formula for the reconstruction of a general 2-D Lambertian object with extent $l(\phi)$.

To specialize this to the case of a finite size line, we set $l(\phi) = l|\cos \phi|$. Equation (A6) becomes

$$f(r, \theta; l) = \int_{-\pi/2}^{\pi/2} \{2\pi^2[l \cos \phi - r \cos(\theta$$
$$- \phi)] - i\pi k\}^{-1} d\phi + \int_{\pi/2}^{3\pi/2} \{2\pi^2[-l \cos \phi$$
$$- r \cos(\theta - \phi)] - i\pi k\}^{-1} d\phi. \tag{A7}$$

We combine these into a single integral. Furthermore we change to Cartesian coordinates with the transformation $x = r \cos \theta$ and $y = r \sin \theta$:

$$f(x, y; l) =$$
$$\int_{-\pi/2}^{\pi/2} \left(\begin{array}{c} \{ik\pi - 2\pi^2[(l + x)\cos \phi + y \sin \phi]\}^{-1} \\ -\{ik\pi + 2\pi^2[(l + x)\cos \phi + y \sin \phi]\}^{-1} \end{array} \right) d\phi. \tag{A8}$$

To obtain the infinitesimal patch response, we find $\lim_{l \to 0}[g(x, y; l)/2l]$, which is

$$f(x, y) =$$
$$\int_{-\pi/2}^{\pi/2} \left[\begin{array}{c} (k - 2\pi i x \cos \phi - 2\pi i y \sin \phi)^{-2} \\ +(k + 2\pi i x \cos \phi + 2\pi i y \sin \phi)^{-2} \end{array} \right] \cos \phi d\phi. \tag{A9}$$

We perform the integration to obtain

To obtain the infinite bandwidth patch response, we find $\lim_{k \to 0} g(x, y)$:

$$\mathrm{PRF}(x, y) = \frac{1}{4\pi^2} \left[\frac{2x}{(x^2 + y^2)^{3/2}} \log \frac{(x^2 + y^2)^{1/2} + x}{(x^2 + y^2)^{1/2} - x} \right.$$
$$\left. - \frac{4}{x^2 + y^2} \right]. \tag{A11}$$

References

1. N. Ahuja and A. L. Abbott, "Active stereo: integrating disparity, vergence, focus, aperture, and calibration for surface estimation," IEEE Trans. Pattern Anal. Mach. Intell. **15,** 1007–1029 (1993).
2. A. Pentland, S. Scherock, T. Darrell, and B. Girod, "Simple range cameras based on focal error," J. Opt. Soc. Am. A **11,** 2925–2934 (1994).
3. S. K. Nayar and Y. Nakagama, "Shape from focus," IEEE Trans. Pattern Anal. Mach. Intell. **16,** 824–831 (1994).
4. S. K. Nayar, M. Watanabe, and M. Noguchi, "Real-time focus range sensor," IEEE Trans. Pattern Anal. Mach. Intell. **18,** 1186–1198 (1996).
5. T. Wilson, ed., *Confocal Microscopy* (Academic, San Diego, Calif., 1990).
6. D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, and J. G. Fujimoto, "Optical coherence tomography," Science **254,** 1178–1181 (1991).
7. J. A. Izatt, M. R. Hee, G. M. Owen, E. A. Swanson, and J. G. Fujimoto, "Optical coherence tomography in scattering media," Opt. Lett. **19,** 590–592 (1994).
8. B. L. Stann, W. C. Ruff, and Z. G. Sztankay, "Intensity-modulated diode laser radar using frequency modulation/continuous wave ranging techniques," Opt. Eng. **35,** 3270–3278 (1996).
9. D. L. Marks, R. A. Stack, D. J. Brady, D. Munson, and R. B. Brady, "Visible cone-beam tomography with a lensless interferometric camera," Science **284,** 2164–2166 (1999).
10. J. Rosen and A. Yariv, "Reconstruction of longitudinal distributed incoherent sources," Opt. Lett. **21,** 1803–1806 (1996).
11. A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging* (Institute of Electrical and Electronics Engineers, New York, 1988).
12. D. I. Marks and D. J. Brady, "Three-dimensional source reconstruction with a scanned pinhole camera," Opt. Lett. **23,** 820–822 (1998).
13. D. L. Marks, R. A. Stack, D. J. Brady, and J. van der Gracht, "Three-dimensional tomography using a cubic-phase plate extended depth-of-field system," Opt. Lett. **24,** 253–255 (1999).

14. M. Born and E. Wolf, *Principles of Optics* (Cambridge U. Press, Cambridge, UK, 1980).

15. E. R. Dowski, Jr. and W. T. Cathey, "Extended depth of field through wave-front coding," Appl. Opt. **34,** 1859–1866 (1995).

16. S. Bradburn, W. T. Cathey, and E. R. Dowski, Jr., "Realizations of focus invariance in optical-digital systems with wave-front coding," Appl. Opt. **36,** 9157–9166 (1997).

17. H. K. Tuy, "An inversion formula for cone-beam tomography," SIAM (Soc. Ind. Appl. Math.) J. Appl. Math. **43,** 546–552 (1983).

18. L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," J. Opt. Soc. Am. A **1,** 612–619 (1984).

19. P. A. Rattey and A. G. Lindgren, "Sampling the 2-D Radon transform," IEEE Trans. Acoust. Speech Signal Process. **ASSP-29,** 994–1002 (1981).

20. A. J. Johnson, "Patch response of cone-beam tomography," M. S. thesis (University of Illinois at Urbana-Champaign, Urbana, Illinois, 1999).

21. A. J. Johnson, D. L. Marks, R. A. Stack, D. J. Brady, and D. C. Munson, Jr., "Three-dimensional surface reconstruction of optical Lambertian objects using cone-beam tomography," in *Proceedings of the IEEE Conference on image processing* (Institute for Electrical and Electronics Engineers, New York, 1999), pp. 663–667.
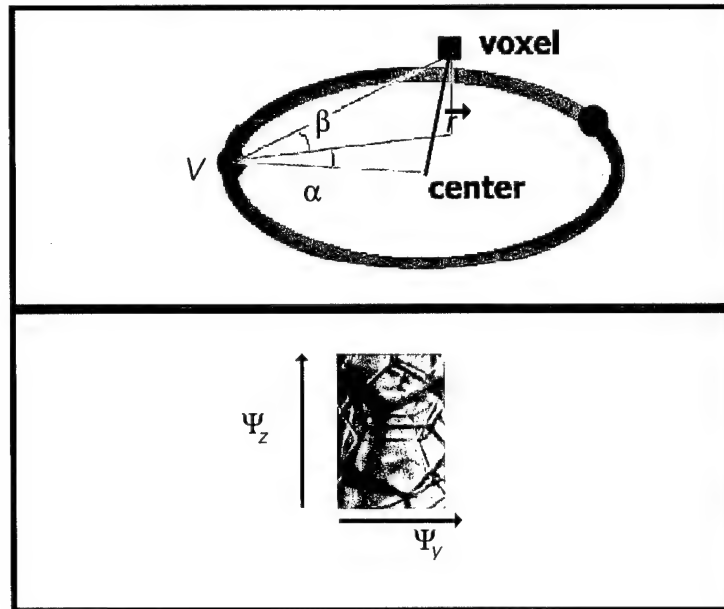
Fig. 3. *Top.* The angular notation used in this paper. Consider a particular voxel and vertex point. We write α for the angle in the vertex plane, β for the angle normal to the vertex plane, and r for the vector that connects the center of the vertex path to the voxel. *Bottom.* Angular notation, continued. For a given image, recorded by a camera, Ψ_y and Ψ_z refer to the coordinates on the camera. It is necessary to find a mapping function between the camera coordinates Ψ_y and Ψ_z, and the points in the reconstructed voxel space, which are denoted by the angles α, β, and the vector r.

faces of the bubbles.

As the object rotates through N steps, an image is taken at each step. Although the object is rotated, one may picture the object as stationary, and the camera as rotating about it Fig. 2 *(bottom)*. Each camera position is referred to as a vertex point P_ϕ, where ϕ describes the angle of the vertex point from the center of the vertex path. The vertex points are all in a circle. The algorithms described here do not require a circular vertex path, but we choose to use one for experimental simplicity. This circle, referred to as the vertex path, lies on the vertex plane. The point V is an arbitrary vertex point, which will be referred to later. The axes are defined such that the x,y, and z axes travel with the camera. The x axis points towards the center of the vertex path [8]. The y axis is in the vertex plane but normal to the vertex path, and the z axis is normal to the vertex plane.

The angular coordinate notation used in this paper is shown in Fig. 3. Consider a certain voxel, and the angles that it makes with respect to a vertex point such as V. We refer to α as the angle in the vertex plane, and β as the angle normal to the vertex plane. The coordinate r connects the center of the vertex path to the voxel. Note that for a given voxel, the values of α and β change, depending on which vertex point we are considering, but the value of r remains constant.

For a given image, recorded by a camera, Ψ_y and Ψ_z refer to the coordinates on the camera. These coordinates are not angles, although each value of Ψ does correspond to an angle projecting out from the camera. It is necessary to find a mapping function between the camera coordinates Ψ_y and Ψ_z, and the points in the reconstructed voxel space, which are denoted by the angles α, β, and a distance coordinate.

Typical images are shown in Fig. 4. Fig. 4 *(left)* shows an image of a test object,

$$\alpha = \arctan \frac{\overrightarrow{r} \cdot \hat{y}'}{d + \overrightarrow{r} \cdot \hat{x}'} \tag{4}$$

$$\beta = \arctan \frac{\overrightarrow{r} \cdot \hat{z}'}{d + \overrightarrow{r} \cdot \hat{x}'} \tag{5}$$

We refer to Eq.4 and Eq.5 as the mapping equations, because they define how the three-dimensional voxel space is to be mapped into the two-dimensional plane of the image. Writing Eq.3 as the tan of the arctan of the angle in Eq.4 may seem somewhat circular, but it is convenient to work with the angle α

Eq.3 may be evaluated in two ways: the voxel oriented method or the pixel oriented method. In the voxel oriented method, which we use in this paper, every voxel is considered. Then, for each voxel, we sum over all pixels. In the pixel oriented method, the rays originating from each pixel are projected through space, and their intersection with the voxel space is calculated. The difference between the voxel-oriented method and the pixel-oriented method is one of computational preference and convenience only, and does not appear in the Feldkamp equations.

3   Distortion Compensation

Our goal was to image the bubbles, which have very fine features. The bubbles must be contained in a cylinder, and this cylinder distorts the light rays. Using ray tracing techniques. we compensate for this distortion, and recover the correct three-dimensional reconstruction. First we discuss the approach to compensating for a generalized distortion, and then we discuss the specific case of the plexiglas cylinder.

3.1   *Distortion compensation for an arbitrary refractive index profile*

In Fig. 5, we illustrate the case of a generalized (2-D) distortion. Assume that we have a known index profile $n(x, y)$ that is completely contained within a circle C with radius R_c. Outside this circle, the index of refraction is $n = 1.0$. In Fig. 5 *(top)*, we show the case when $n(x, y) = 1$. Then we can connect the vertex point V and a voxel P with a straight line, and the equations described in the previous section apply.

Next we consider the situation when $n(x, y)$ is a known function, as shown in Fig. 5 *(bottom)*. Here, the blue circle represents a region where $n(x, y)$ is modified (e.g. $n(x, y) = 2$ in this region). We cannot compensate for this distortion by simply stretching the image. The reason is that distorted rays do not follow the same path as any of the undistorted rays. That is, the rays in the undistorted space contain a different set of voxel points than the rays in the distorted space. Therefore, no matter how the image is mapped, these rays will not coincide. We note that, for the special case of the cylinder, it may be possible to correct the reconstruction by shifting the apparent position of the vertex point and stretching the images, since the cylinder acts as a lens. However, we chose the approach described here because it applies to a more general class of arbitrary distortions $n(x, y)$, and it is also more exact.

Under the distortion, the equations Eq.1, Eq.2, and Eq.3 will still be valid. However, we must change the mapping functions defined in Eq.4 and Eq.5. First, we will find the ray that connects P and V. The point P' is defined as the intersection of this ray with the circle C. The angle that is necessary for the mapping functions (in Eq.4 and Eq.5.) is the angle between P' and V.

The problem of connecting 2 points by ray-tracing, given an arbitrary refractive index profile, is solved by Fermat's principle of least time [12]. To solve it numerically, we require that Snell's Law [12] is satisfied at all interfaces, and that the ray intersects P and V. This inverse problem can be approached with a simple searching algorithm.
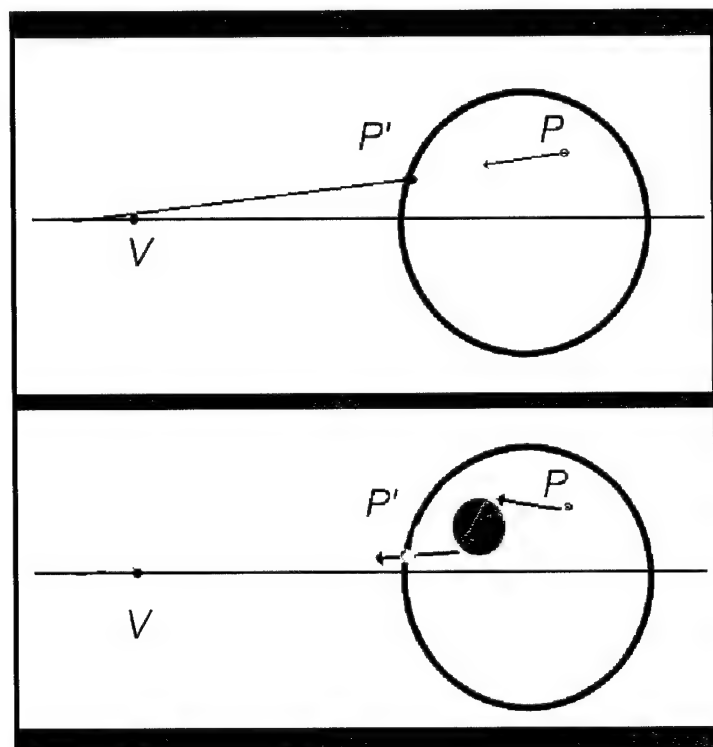
Fig. 5. Compensating for distortion. *Top.* This illustrates the case of a generalized (2-D) distortion. Assume that we have a known index profile $n(x, y)$ that is completely contained within a circle C with radius R_c. Outside this circle, the index of refraction is $n = 1.0$. Here, we show the case when $n(x, y) = 1$. Then we can connect the vertex point V and a voxel P with a straight line, and the equations described in section 2 apply. *Bottom.* Consider the situation when $n(x, y)$ is a known function. In this case, the blue circle could represent a region where $n(x, y) = 2$; everywhere else, $n(x, y) = 1$. Using Snell's Law and numerical iteration, we find the ray that connects P to V. Starting at a voxel P, and given a direction vector, we find the intersection of the ray with the circle C at P', Then the angles (for Eq.4 and Eq.5) can be found from the points M and P'.

Fig. 6. This shows one step in the construction to find the ray that connects the voxel to the vertex point. Note that this is a two-dimensional calculation. In this step, we project a ray from the voxel towards the vertex point. Using Snell's law, we find the angle of refraction that occurs when the ray intersects the inner radius of the cylinder. Not shown is the next step, in which we calculate the refraction of the ray at the outer air-cylinder interface.
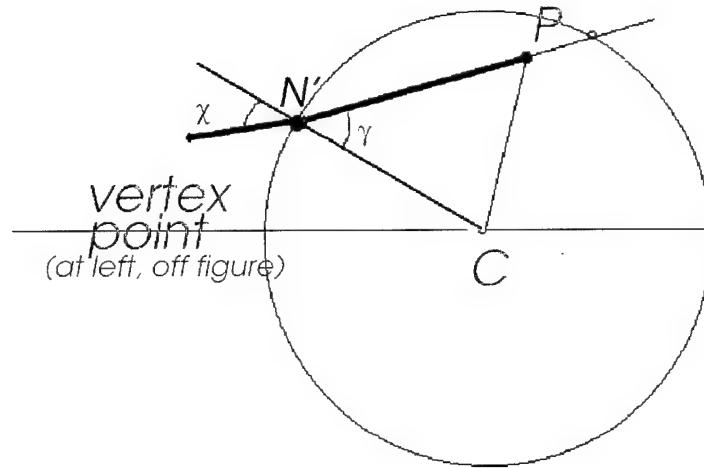
We start from the voxel, P, and assume a direction vector. We shoot a trial ray into the volume, and measure the distance between the ray and the vertex point V. Then we iterate the initial direction vector until the ray intersects the vertex point.

3.2 *Distortion compensation for the specific case of the plexiglas cylinder*

The plexiglas cylinder is easy to model because it is an exact circle in the $x - y$ plane. Thus, we can solve for the exact angles, rather than modeling the refractive index profile on a grid. Assume that the cylinder has inner radius R_1, outer radius R_2, and index of refraction n. Because the cylinder has no curvature in the plane normal to the vertex plane, Eq.5 remains unchanged. Eq.4 must be changed because the distortion will alter the angle α. This is now a two-dimensional problem.

Fig. 6 is a diagram in which we show one step in the construction used to solve for the ray path. This diagram represents the general case in which we have a voxel with position $\vec{r}$ inside a circle, and a direction vector. We wish to find the point N' at which this ray will intersect the circle, as well as the output direction vector. The point N' is found by following the initial direction vector until the circle is intersected. The radius connecting N' with the center of the circle C makes a 90° angle with the tangent to the circle. Thus, we can find the angle γ, and then using Snell's Law, find the angle χ. In this case we take the inner index of refraction as n_1 and the outer index of refraction as n_2. Note that we are considering a cylinder with a given thickness, so that we will repeat this calculation twice. The first time, we will assume that the cylinder has $n_1 = 1.0$ (air) inside, and $n_2 = 1.49$ (plexiglas) outside. The second time, after finding the point N', we will then assume that the cylinder has plexiglas inside and air outside. The point at which this ray intersects the x-axis is then found.

In Fig. 7, we show a ray tracing diagram with several rays plotted. This represents the result of the calculation described above, as well as the searching algorithm to find
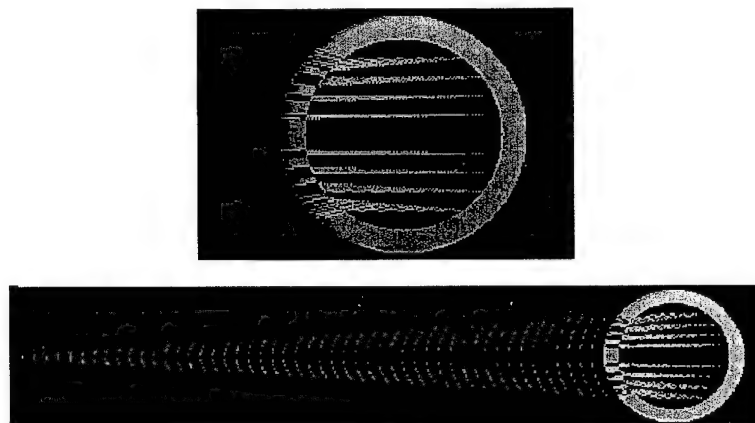
Fig. 7. Solving for the rays such that they intersect the vertex point. *Top.* Magnified view of cylinder. *Bottom.* This image shows the cylinder as well as the vertex point.
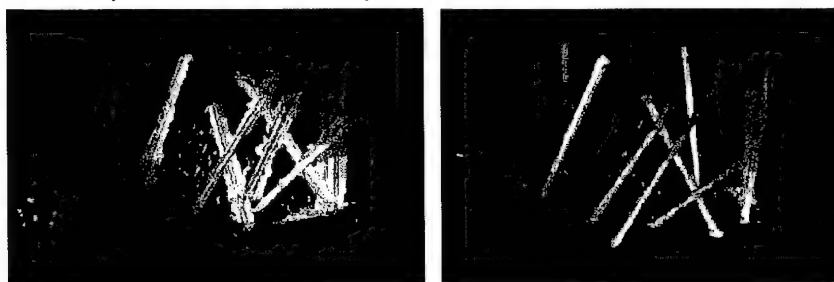

Fig. 8. *Left.* Reconstructing the test object of the needles without applying the distortion algorithm. This three-dimensional image was generated by vtk[10]. The colors in this picture are an arbitrary colormap and have no significance. *Right.* The image is improved through application of the distortion algorithm. The red area in the back of this image is a piece of paper that was in the original object.

the correct direction vector that will intersect the vertex point.

4   Results and Analysis

In Fig. 8 (*left*), we show the three-dimensional reconstruction of the test object from Fig. 4 (*left*). This reconstruction is done without correcting for the distortion. It can be seen that the features are somewhat blurry. Fig. 8 (*right*) shows the substantial improvement when the correction algorithm is applied. Fig. 9 (*left*) is a slice of the dataset (without distortion correction) that is normal to the $\hat{z}$ axis, and Fig. 9 (*right*) is a similar slice of the dataset (with distortion correction). The distortion will be more significant for points that are closer to the edge of the cylinder. As shown in Fig. 7, the rays from such edge points are modified more than points that are closer to the center. This effect can be seen in Fig. 9 (*left*), where points farther away from the center appear as blurred crosses, but points closer to the center of the cylinder appear as sharper points.

We then applied this algorithm to the case of the bubbles, using the input data shown in Fig. 4 (*right*). Fig. 10 (*left*:uncorrected and *right*:corrected) shows the result with and without the distortion correction algorithm, Clearly, after applying the correction algorithm, the image is substantially improved. Slices of the bubble dataset are shown
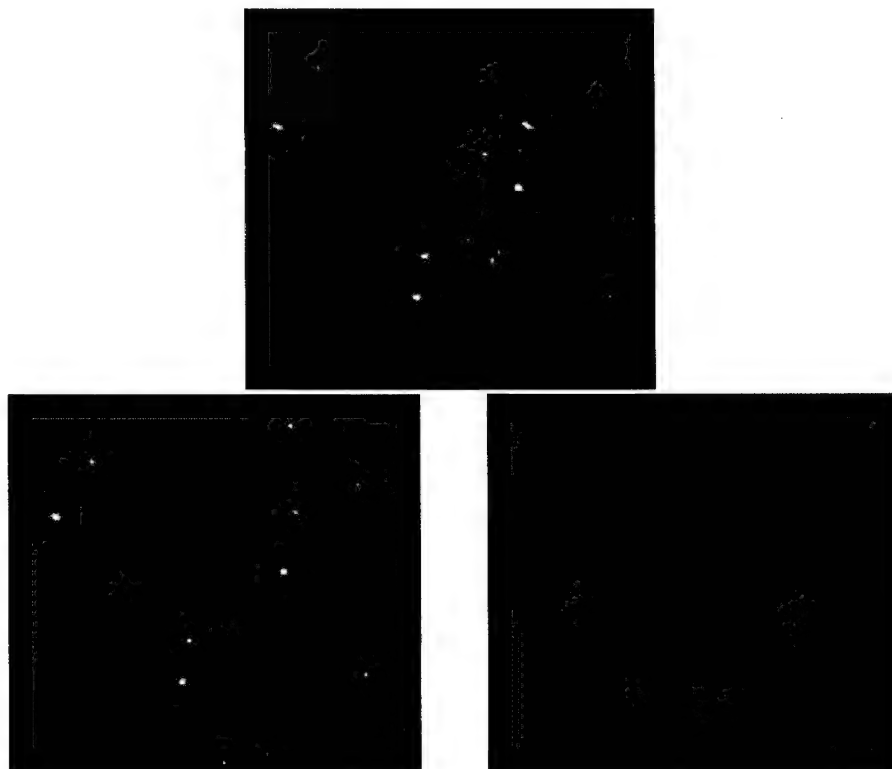
Fig. 11. *Top.*This is a slice through the dataset of Fig. 10, which is a reconstruction of the bubbles without applying the distortion correction. There is some blurring in this image. *Bottom Left.* The distortion correction algorithm is applied, improving the bubbles images. *Bottom Right.* An erosion algorithm is applied to reduce each blob in the image at left (the corrected bubbles images) to a single point.

Fig. 9. *Left.* This is a slice through the dataset of Fig. 8, which is a reconstruction of the needles test object without applying the distortion correction. The needles, which should appear as points, in this case appear as blobs. Some of the features appear as crosses. *Right.* The distortion correction algorithm is applied. The cross section of the needles now appear as points.
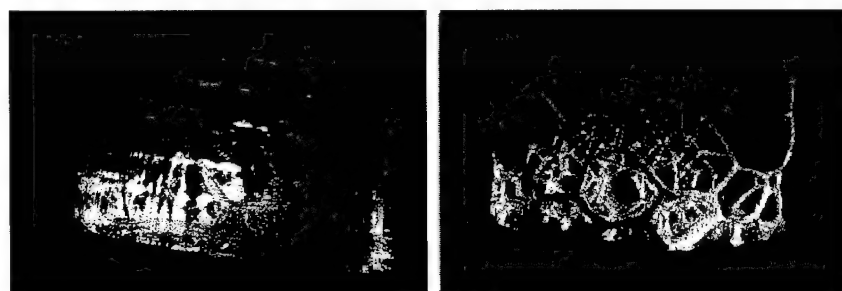
Fig. 10. *Left.* Reconstructing the bubbles without applying the distortion algorithm. Only a slice of the image from Fig. 4 is reconstructed. *Right.* The image is improved through application of the distortion algorithm.

in Fig. 11 (*top*:uncorrected and *bottom left*:corrected).As in the example of the needles, the closer the points are to the center of the cylinder (in the uncorrected image), the less they are distorted. For studying the foam, it will be important to have a complete set of data that includes the points closest to the cylinder edge, so that this correction is necessary. The improvement with this correction algorithm seems quite clear, although the corrected image still shows some blurring for points close to the edge of the cylinder. The image may be further improved with image processing. In Fig. 11 (*bottom right*), we show an erosion [13] algorithm applied to the data of Fig. 11 (*bottom left*).

5 Conclusions

In this paper we show reconstruction of a three-dimensional foam. The foam is reconstructed using tomography algorithms. Using an algorithm that incorporates ray tracing, we are able to compensate for the distortion induced by a plexiglas cylinder. It is shown that this algorithm improves the images for the case of a test object, as well as for the bubbles. This distortion correction algorithm may be useful in various areas of tomography.

One area of future work will include improving the initial images. This could include redesign of the container, with thinner walls, to reduce or eliminate the distortion problem. Although this would improve our foam imaging, it is noted that part of our goal was to study the computational correction of optical distortion. The illumination of the Plateau borders could also be improved. In [14], the authors dissolve fluorescein salt in their foaming solution and illuminate with ultra-violet light. With this technique, the Plateau borders fluoresce and there is no stray light.

We are currently analyzing the 3-dimensional data set to reveal the exact polyhedral configuration and its evolution. This work includes signal processing techniques such as matched filtering.

Three-dimensional tomography using a cubic-phase plate extended depth-of-field system

Daniel L. Marks, Ronald A. Stack, and David J. Brady

Department of Electrical and Computer Engineering, Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

Joseph van der Gracht

Army Research Laboratory, AMSRL-SE-EO, 2800 Powder Mill Road, Adelphi, Maryland, 20783

We use cubic-phase plate imaging to demonstrate an order-of-magnitude improvement in the transverse resolution of three-dimensional objects reconstructed by extended depth-of-field tomography. Our algorithm compensates for the range shear of the cubic-phase approach and uses camera rotation to center the reconstructed volume on a target object point. © 1999 Optical Society of America
  OCIS codes: 110.6880, 110.6960, 110.4850, 220.1230, 100.1830.

Inversion of the line integrals associated with extended depth-of-field (EDF) imaging has been used in the x-ray regime to reconstruct three-dimensional (3D) objects.[1-3] Related ray-projection techniques have been used in the visible range to reconstruct radiant sources.[4] In Ref. 5 we used EDF pinhole imaging to reconstruct a 3D volume in the visible spectral range. The relatively poor transverse resolution of the pinhole camera is a shortcoming for this technique. The cubic-phase plate (CPP) EDF system[6-9] measures similar line integrals while maintaining a relatively large system aperture. The larger aperture yields superior transverse resolution and light-gathering efficiency. In this Letter we describe the use of a CPP EDF system to form 3D images.

Aperture modulation in CPP imaging yields a point spread function (PSF) that is relatively insensitive to defocus. Digital deconvolution of the PSF yields an image focused at all depths for which the range invariance holds. The transmittance of a CPP is $t(x,y) = \exp[j\alpha(x^3 + y^3)]$.[6] Assuming an unbounded aperture is the Airy Ai function, and γ is the defocus parameter $(1/z + 1/z' - 1/f)$, the PSF for quasi-monochromatic imaging between a source point at (x,y,z) and an image point at (x',y',z') can be shown to be

$$\text{PSF}\left(x' + \frac{z'}{z}x, y' + \frac{z'}{z}y, z, z'\right)$$

$$= \frac{1}{\alpha^{4/3}} Ai^2\left\{-\frac{\alpha^{2/3}}{3^{1/3}}\left[\frac{2\pi}{\lambda z'\alpha}\left(x' + \frac{z'}{z}x\right) + \frac{\pi^2\gamma^2}{3\alpha^2\lambda^2}\right]\right\}$$

$$\times Ai^2\left\{-\frac{\alpha^{2/3}}{3^{1/3}}\left[\frac{2\pi}{\lambda z'\alpha}\left(y' + \frac{z'}{z}y\right) + \frac{\pi^2\gamma^2}{3\alpha^2\lambda^2}\right]\right\}, \quad (1)$$

where $Ai(x)$ is the Airy Ai function[10] and f is the focal length. To form a range-independent deblurred image, we deconvolve the PSF from the detected image. The deblurred image is sheared with respect to range, but the shear can be removed in 3D imaging. In practice, the finite plate aperture limits the range independence of the PSF. The in-

finite aperture approximation assumes that the cubic phase varies more rapidly than the natural quadratic phase at the aperture edge. This criterion implies that γ must be comparable with $\alpha d\lambda$, where d is the system aperture.

We consider the CPP EDF camera imaging source volume in Fig. 1. The camera acquires a series of images as it is translated laterally to the optical axis. The lateral position of the center of the camera principal plane is denoted $\hat{x}$. Unlike in Ref. 5, the camera axis rotates about a point in the source space during lateral displacement. The rotation centers the reconstructed range field on the pivot point. The nominal distance from the pivot point to the center of the camera aperture is z_0. We form 3D images by capturing a series of distorted 2D images for various values of $\hat{x}$, digitally deconvolving each frame to obtain an in-focus EDF 2D image and transforming the series of EDF images to obtain a 3D source model.

Assuming that the pivot point satisfies the imaging condition, $z_0 = z'f/z' - f$, and defining $\Delta z = z - z_0$, the origin of the PSF in the (x',y') plane for the source point at (x,y,z) is

$$\frac{x'}{z'} = -\frac{x}{z} - \frac{\pi}{6\alpha\lambda z_0^2}\left(\frac{\Delta z}{z}\right)^2,$$

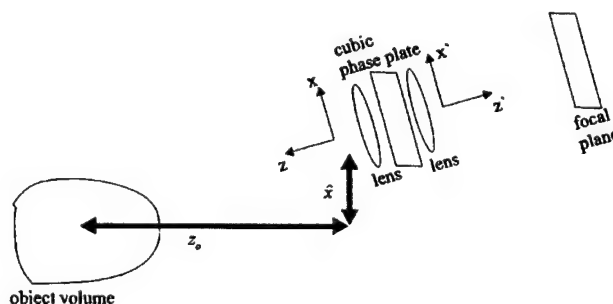$$\frac{y'}{z'} = -\frac{y}{z} - \frac{\pi}{6\alpha\lambda z_0^2}\left(\frac{\Delta z}{z}\right)^2.$$

Fig. 1.   System geometry.

© 1999 Optical Society of America

For a given value of $\hat{x}$ the deconvolved image is approximated by

$$I(\xi, \eta) = \int P(x, y, z)\delta\left[\frac{\xi}{z'} + \frac{(x + \hat{x})z_0}{z} - \frac{\hat{x}}{z_0}\right.$$
$$+ \frac{\pi}{6\alpha\lambda z_0^2}\left(1 - \frac{z_0}{z}\right)^2\right]\delta\left[\frac{\eta}{z'} + \frac{y}{z}\right.$$
$$+ \frac{\pi}{6\alpha\lambda z_0^2}\left(1 - \frac{z_0}{z}\right)^2\right]dx\,dy\,dz, \qquad (2)$$

where $P(x, y, z)$ is the source power density. Inverting Eq. (2) as in Ref. 5, we Fourier transform with respect to (ξ, η) and introduce the variables $x_p = (z'/z)x$, $y_p = (z'/z)y$, $q = k_\xi z'\hat{x}$, and $z_p = 1/z$ to obtain

$$\tilde{I}(k_\xi, k_\eta, q) = \int P(x_p, y_p, z_p)\exp(jk_\xi x_p)\exp(jk_\eta y_p)$$
$$\times \exp(jqz_p)\exp(-jqz_{0p})\exp\left[jk_\xi\frac{\pi z'}{6\alpha\lambda z_0^2}\right.$$
$$\times (1 - z_0 z_p)^2\right]\exp\left[jk_\eta\frac{\pi z'}{6\alpha\lambda z_0^2}\right.$$
$$\times (1 - z_0 z_p)^2\right]dx_p\,dy_p\,dz_p. \qquad (3)$$

Fourier inversion of Eq. (3) yields the focused 3D source distribution evaluated at

$$\left[x_p + \frac{\pi z'}{6\alpha\lambda z_0^2}(1 - z_0 z_p)^2,\right.$$
$$\left.y_p + \frac{\pi z'}{6\alpha\lambda z_0^2}(1 - z_0 z_p)^2, z_p - \frac{1}{z_0}\right],$$

where (x_p, y_p, z_p) are the Fourier conjugate variables for (k_ξ, k_η, q). The range in (k_ξ, k_η, q) over which $\tilde{I}(k_\xi, k_\eta, q)$ can be sampled defines the 3D bandpass, or band volume, for the imaging system. The 3D Fourier transform of the band volume is the system impulse response. Since the longitudinal Fourier sampling coordinate, q, is proportional to the transverse sampling coordinate, k_ξ, there is a missing cone in the band volume.[11] The longitudinal resolution for low-transverse-frequency objects is limited as a result of the missing cone.

We confirmed our model for 3D imaging by experimentally measuring the impulse response. Our experimental system consisted of two 20-cm focal-length lenses (yielding a nearly 10-cm effective focal length) separated by a 1.2-cm-aperture CPP pupil with $\alpha = 58.6$ cm^{-3}. We placed the sources and the focal-plane array approximately 20-cm from the CPP to achieve nominal 1:1 imaging. The focal-plane array was a 512 by 512, 1.27-cm-square Princeton Instruments back-illuminated 16-bit-resolution CCD camera. The camera and the imaging system were attached to metal rods and placed on a computer-controlled Newport rotation stage, which was attached to an Aerotech computer-controlled translation stage with 5-cm travel.

A 660-nm laser diode that operated as a LED served as the point source for building the digital deconvo-

lution filter. We placed the source 20 cm from the cubic-phase mask and sensed the raw PSF at the focal plane. Because the cubic-phase transmittance is rectangularly separable, we sampled the PSF on the x and y axes and computed an inverse weighted Wiener filter of the sampled data to obtain separate vertical and horizontal deconvolution filters. We selected the filter weight to band limit the inverse filter to prevent noise from dominating the image. We used ~50% of the bandwidth of the system. The digitally deconvolved PSF occupied from 1.5 to 2 camera pixels, and the raw PSF covered approximately 15 pixels. The CCD camera's pixels were 22 μm square.

We acquired frames by translating the imaging system at fixed intervals. At each acquisition position we rotated the camera toward the source pivot point and integrated the intensity on the focal plane for 30 ms. The total length of travel of the translation stage was 4.5 cm, and the total rotation angle was ~13°. 256 images were taken over this path. The projections were sampled at 512 by 512 pixels and deconvolved. After deconvolution, we resampled the projections to compensate for the camera rotation and to compress the frame size to 256 by 256. The resampled images corresponded to projections displaced linearly in the tangents of the angles as measured from each projection origin. We combined the projections into a 3D model of the source by computing the 2D Fourier transform of each image, resampling the k_ξ axis onto a Cartesian grid, and then taking the 3D Fourier transform. These operations were performed in $O(N^3 \log N)$ time.

We tested our system by reconstructing the 660-nm point source at approximate distances of 15, 17, 20, and 25 cm from the principal plane of the imaging system. Figure 2 shows transverse slices through each of the reconstructed 3D PSF's. The axes correspond to projection angle. Each axis spans 52 mrad. The 15-cm image is significantly worse because the sharp increase in defocus as one approaches the principal plane causes the invariant raw PSF approximation to fail. Longitudinal slices through the reconstructed 3D PSF's are shown in Fig. 3. The horizontal axis
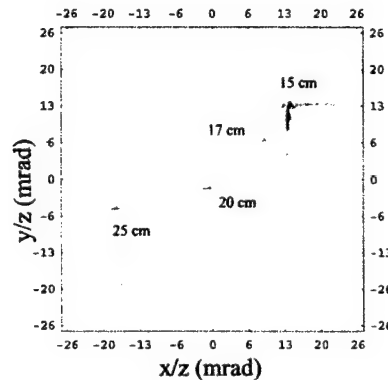

Fig. 2. Superimposed lateral cross sections of four digitally reconstructed 3D PSF's, labeled with their distances from the principal plane. The axes are angles labeled in milliradians.
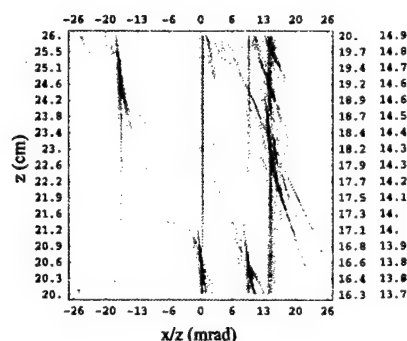
Fig. 3. Superimposed longitudinal cross sections of four digitally reconstructed 3D PSF's labeled with their distances from the principal plane. The horizontal axis is angle labeled in milliradians, and the vertical axis is projective depth labeled in centimeters. The vertical scale is plotted linearly in $1/z$ space but is marked in z space.
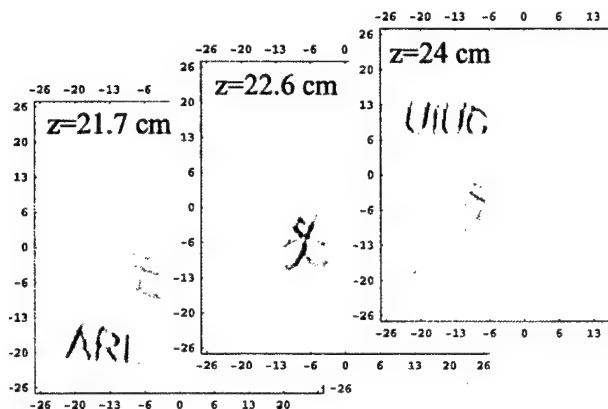

Fig. 4. Three lateral slices through the reconstruction of a demonstration source. The axes correspond to angular position relative to the principal axis in milliradians.

represents the transverse resolution in milliradians, and the vertical axis represents the projective depth in inverse meters. The vertical axis spans 1.16 m^{-1}. The projective depth is the depth that is linear in z^{-1} and not in z. The vertical axes are marked in z coordinates, however. Three separate axes are shown because the reconstructed sources are aliased onto the plot from three separate reconstruction patches. The reconstruction uses a projective depth step size of 0.0045 m^{-1}, which is determined by the total linear length of travel of the camera ($\hat{x}_{\text{span}} = 4.5$ cm) and the angular resolution of the camera ($\theta_{\text{res}} = 0.2$ mrad) by $z_{\text{res}}' = \theta_{\text{res}}/\hat{x}_{\text{span}}$. The achieved resolution is limited by incomplete coverage of the 3D Fourier space and the band limit in the deconvolution kernel. The measured longitudinal sizes of the source are 0.216 m^{-1} (4.9 mm) for the 15-cm trial, 0.153 m^{-1} (4.4 mm) for the 17-cm trial, 0.122 m^{-1} (4.8 mm) for the 20-cm trial, and 0.189 m^{-1} (11.8 mm) for the 25-cm trial. As indicated by the 17- and the 25-cm reconstructions, the depth-of-field exceeds 1 m^{-1}. This range is consistent with the value of $\alpha d \lambda = 0.46 \text{ m}^{-1}$

for our system. The pinhole size needed to achieve this depth of field at 660 nm is 800 μm. The transverse resolution for the pinhole system would be approximately one order of magnitude worse than the $80-100$-μm resolution indicated for the cubic-phase system in Fig. 2.

Figure 4 demonstrates 3D reconstruction of a complex source. The source consisted of white-on-black text and images on small strips of paper. The papers were illuminated by two broadband fluorescent lamps to provide uniform illumination. The paper in the front had the letters ARL on it, the paper in the middle had a stick figure likeness of a man, and the paper in the rear had the letters UIUC on it. The exposed sections of the papers were approximately 0.3 by 0.4 cm in size. Figure 4 shows three cross sections of the source at various depths. The intensity of the source is shown in reverse, and the negative intensity artifacts that are due to the PSF have been filtered out. Although the reconstruction is fairly accurate on these planes, it should be noted that the PSF spreads the images over many planes, as can be seen because the stick figure man in the center plane appears weakly in the other two planes.

The cubic-phase plate extended depth-of-field system yields substantially better transverse resolution than a pinhole with comparable depth of field. Since the transverse resolution is coupled to longitudinal resolution through the q variable, increased transverse resolution translates directly into improved longitudinal resolution for a given scan range. Further investigation into the optimality of the cubic-phase modulation, the effect of noise in deconvolution, and deconvolution filter design is required for characterization and exploitation of this improvement in longitudinal resolution.

References

1. Y. W. Chen, N. Miyanaga, and N. Yamanaka, J. Appl. Phys. **68**, 1483 (1990).
2. J. W. V. Gissen, M. A. Viergever, and C. Graaf, IEEE Trans. Med. Imaging **MI-4**, 91 (1985).
3. L. I. Yin and S. M. Seltzer, Appl. Opt. **32**, 3726 (1993).
4. I. Ashdown, J. Illum. Eng. Soc. **22**, 163 (1993).
5. D. Marks and D. Brady, Opt. Lett. **23**, 820 (1998).
6. E. Dowski and W. Cathey, Appl. Opt. **34**, 1859 (1995).
7. J. van der Gracht, E. Dowski, and W. Cathey, Proc. SPIE **2537**, 279 (1995).
8. J. van der Gracht, E. Dowski, M. G. Taylor, and D. M. Deaver, Opt. Lett. **21**, 919 (1996).
9. S. Bradburn, W. Cathey, and E. Dowski, Appl. Opt. **36**, 9157 (1997).
10. W. F. Magnus, F. Oberhettinger, and R. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Spinger-Verlag, New York, 1966), p. 76.
11. M. Y. Chiu, H. H. Barrett, R. G. Simpson, C. Chou, J. W. Ardent, and G. R. Gindi, J. Opt. Soc. Am. **69**, 1323 (1979).

Efficient Source State Estimation

Confocal microscopy with a volume holographic filter

George Barbastathis* and Michal Balberg

*Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana–Champaign,·
405 North Mathews Avenue, Urbana, Illinois 61801*

David J. Brady

*Beckman Institute for Advanced Science and Technology and Department of Electrical and Computer Engineering,
University of Illinois at Urbana–Champaign, 405 North Mathews Avenue, Urbana, Illinois 61801*

We describe a modified confocal microscope in which depth discrimination results from matched filtering by a volume hologram instead of a pinhole filter. The depth resolution depends on the numerical aperture of the objective lens and the thickness of the hologram, and the dynamic range is determined by the diffraction efficiency. We calculate the depth response of the volume holographic confocal microscope, verify it experimentally, and present the scanned image of a silicon wafer with microfabricated surface structures. © 1999 Optical Society of America

OCIS codes: 180.1790, 110.6880, 090.7330.

The pinhole preceding the detector in a confocal microscope is a shift-variant optical element. On-axis in-focus point-source objects are imaged exactly inside the pinhole and give maximal intensity. An out-of-focus object, even when it is on axis, is equivalent to an extended source on the input focal plane. The off-axis portion of this extended source is filtered out by the limited aperture of the pinhole. Theoretically, the depth resolution is optimal when an infinitesimally small pinhole is used.[1] However, such a device is an *ad hoc* filter that does not match perfectly the impulse response of any realistic optical system. In practice, the minimum pinhole size, and hence the depth-resolution limit, are determined by light efficiency (i.e., the required dynamic range of the measurement) and the broadening of the focal spot by lens aberrations.[2] Coupling the dependence of two functional requirements (depth resolution and dynamic range) to a single design parameter (the pinhole size) is a poor design choice.[3] This is evident when the collected light has low intensity, e.g., in fluorescence and two-photon confocal microscopy.

In this Letter we present a new confocal imaging principle in which the pinhole is replaced with a matched filter recorded on a volume hologram. The hologram is recorded such that the field that is generated by an in-focus object is maximally diffracted, whereas objects that are out of focus are filtered out because they are Bragg mismatched. Consequently, dynamic range and axial resolution are decoupled; the dynamic range is determined by the diffraction efficiency of the volume hologram, and the axial resolution by the numerical aperture of the objective lens and the thickness of the hologram. Additional benefits of pinhole-free confocal microscopy are ease of alignment and improved aberration performance: Objective-lens aberrations are phase conjugated out during the hologram reconstruction process, and collector-lens aberrations (which increase the collected spot size) are irrelevant in the absence of a pinhole.

The volume holographic confocal microscope is shown schematically in Fig. 1. The volume hologram is recorded by the interference of two coherent beams at wavelength λ. The objective lens brings the first beam to focus on a reference surface, one focal distance F away from the objective. The reflected beam is recollimated by the same objective and is used as the recording plane-wave reference beam, with wave vector $\mathbf{k}_R = (2\pi/\lambda)\hat{z}$. The signal beam is a plane wave that is incident upon the recording medium along $\mathbf{k}_S = (2\pi/\lambda)\hat{x}$ (90° recording geometry). The resulting grating vector is $\mathbf{K} = \mathbf{k}_S - \mathbf{k}_R$. During the imaging operation, the signal beam is blocked. The reference surface is replaced by the object surface, and the reflected beam reconstructs the volume hologram. The diffracted light is collected by a second objective lens (focal length F') and captured by a photodetector.

Compared with a reflection-mode confocal microscope, the imaging arrangement shown in Fig. 1 contains two modifications, in addition to the volume hologram: (a) the objective lens is placed in a Fourier-transform rather than an imaging configuration and (b) the aperture in front of the detector does not contribute to depth discrimination but only limits scatter and other light-noise sources. If the reconstructing object is in focus (dotted lines in Fig. 1), this device
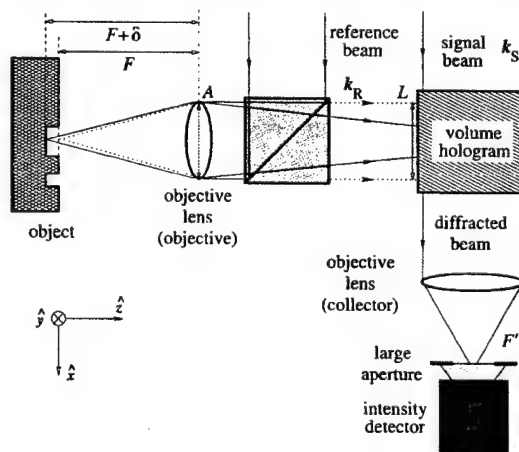

Fig. 1.   Volume holographic confocal microscope without a pinhole at the detector plane.

operates exactly like a confocal microscope, because the volume hologram is Bragg matched (the recording and the reconstructing reference beams are identical); therefore the diffracted intensity reaching the detector is maximum.

Consider now an object that is defocused by a small distance δ. The beam that is reflected from the object is no longer collimated by the objective lens but contains an angular spectrum of plane-wave components, as shown by the solid lines in Fig. 1. Diffraction of the off-axis components by the volume hologram is weaker because of Bragg mismatch. Consider the component with wave vector $\mathbf{k}_p = (2\pi/\lambda)\{u\hat{x} - v\hat{y} + [1 - (u^2 + v^2)/2]\hat{z}\}$, shown in Fig. 2 ($|u|, |v| \ll 1$). Born's first approximation in volume diffraction theory[4] requires that the diffracted wave vector $\mathbf{k}_d$ have the same $\hat{y}$ and $\hat{z}$ and components as the vector $\mathbf{k}' = \mathbf{k}_p + \mathbf{K}$ and, moreover, that $|\mathbf{k}_d| = 2\pi/\lambda$; therefore

$$\mathbf{k}_d = \frac{2\pi}{\lambda}\left[\left(1 - \frac{v^2}{2}\right)\hat{x} - v\hat{y} - \frac{u^2 + v^2}{2}\hat{z}\right]. \quad (1)$$

Taking only one diffracted component, $\mathbf{k}_d$, into account in effect neglects the finite extent of the hologram in the $\hat{y}$ and $\hat{z}$ dimensions. However, the analysis remains valid because the entire spatial spectrum that is diffracted in response to $\mathbf{k}_p$ behaves (in the paraxial approximation) similarly to its central plane-wave component $\mathbf{k}_d$, which is the only component that we consider here. In other words, the impulse response that is due to the finite hologram aperture does not affect the depth discrimination of the system.

The diffracted intensity along this central component $\mathbf{k}_d$ is proportional to $\mathrm{sinc}^2(\Delta k_x L/2\pi)$, where L is the extent of the hologram in the $\hat{x}$ direction, and $\mathrm{sinc}(\xi) \equiv \sin(\pi\xi)/(\pi\xi)$. The quantity Δk_x is the deviation of $\mathbf{k}'$ from the k sphere (see Fig. 2):

$$\Delta k_x = |\mathbf{k}' - \mathbf{k}_d| = \frac{2\pi}{\lambda}\left(u + \frac{v^2}{2}\right). \quad (2)$$

To obtain the overall diffraction efficiency summed over an infinite detector area we integrate the diffracted intensities from all spatial frequency components $\mathbf{k}_p$ that are allowed through the circular objective aperture (diameter A; Fig. 1) and normalize them for a total incident power of 1. The result is

$$\eta(\delta) = \frac{\eta_0}{\pi} \int_0^{2\pi} d\theta \int_0^1 d\rho\, \rho\, \mathrm{sinc}^2\left\{(\mathrm{NA})^2|\delta|\rho\, \frac{2L}{\lambda A}\right.$$

$$\left. \times \left[\cos\theta + (\mathrm{NA})^2|\delta|\rho\, \frac{\sin^2\theta}{2A}\right]\right\}, \quad (3)$$

where $\eta_0 \equiv \eta(0)$, $(\mathrm{NA}) \approx A/(2F)$ is the objective numerical aperture, $|\delta|/F \ll 1$ is assumed, and polar coordinates (ρ, θ) are substituted for (u, v) in the integral. A microscope without a pinhole in front of the detector corresponds to the case $L = 0$, when the total detected intensity does not depend on object depth. For finite thickness $L > 0$, the integral increases with $|\delta|$ much slower than the denominator δ^2, and the dif-

fracted intensity decreases rapidly as a result. The instrument is optimal if all the light coming out of the objective reaches the hologram, i.e., $L = A$.

The Bragg-mismatch effect (expressed through the sinc function in the integrand) effectively acts as a matched spatial filter, discarding the defocused light. This shift-variant filtering operation is similar to the field-of-view limitation imposed by the pinhole of a confocal microscope. The passband has an elliptical shape, with semiaxes $u_{max} = \lambda/L$ and $v_{max} = \sqrt{2\lambda/L}$. Since $v_{max} \gg u_{max}$, the depth response is determined primarily by the term $(\mathrm{NA})^2|\delta|\rho$ in the argument of the sinc function of Eq. (3). As a measure of depth resolution, we use the FWHM of $\eta(\delta)$. By fitting numerical data from Eq. (3) at the optimal geometry $L = A$, we obtain

$$\delta_{\mathrm{FWHM}} = 1.09 \times \frac{\lambda}{(\mathrm{NA})^2}. \quad (4)$$

By comparison, a confocal microscope with zero pinhole size has $\delta_{\mathrm{FWHM}} = 0.86 \times \lambda/(\mathrm{NA})^2$, but the FWHM increases rapidly with pinhole size in realistic systems.[1]

We implemented the pinhole-free confocal microscope shown in Fig. 1 experimentally. We used an Ar^+ laser ($\lambda = 488$ nm) as a light source; a 1-cm^3 $\mathrm{LiNbO_3}$:Fe crystal (45° cut; refractive index, ≈ 2.2) as a holographic medium; a 60×, NA 0.85 objective lens ($A \approx 5$ mm); and a 10×, NA 0.25 collector lens. The reference and the object surfaces were polished silicon wafers with microfabricated features, mounted upon a Klinger translation stage (0.1-μm step size) with three degrees of freedom. The light collected through a variable aperture was measured with a UDT photodetector. To implement a confocal microscope in the same experimental arrangement we simply replaced the volume hologram with a mirror oriented at 45°, directing the reflected beam into the collector lens.

The dependence of the normalized diffraction efficiency on the depth of the object surface is shown by curves (a) and (b) of Fig. 3. The depth resolution is the same for aperture sizes of 25 μm (matched to the
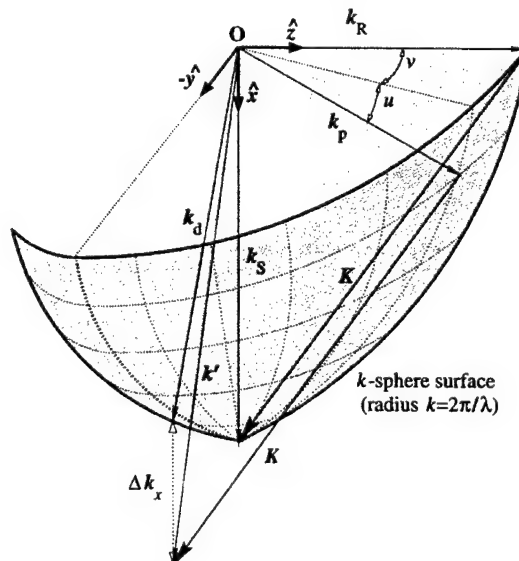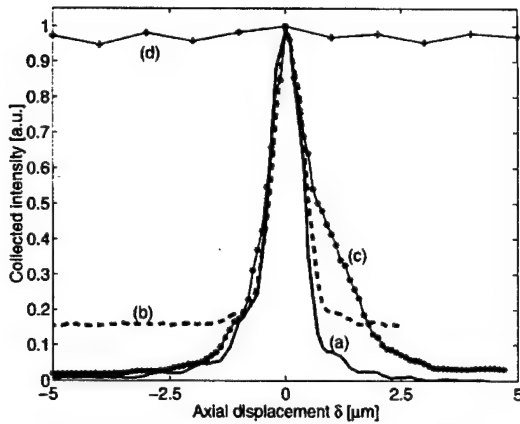

Fig. 2.   Bragg mismatch in the k sphere.

Fig. 3. Collected intensity as a function of object depth δ for the volume holographic confocal microscope with (a) 25-μm and (b) 1-mm pinholes and for the confocal microscope (with a 45°-oriented mirror replacing the volume hologram) with (c) 25-μm (d) 1-mm pinholes. Location $\delta = 0$ corresponds to the depth of the reference surface (at the focal plane of the objective lens). All curves are normalized such that their peak values equal 1.
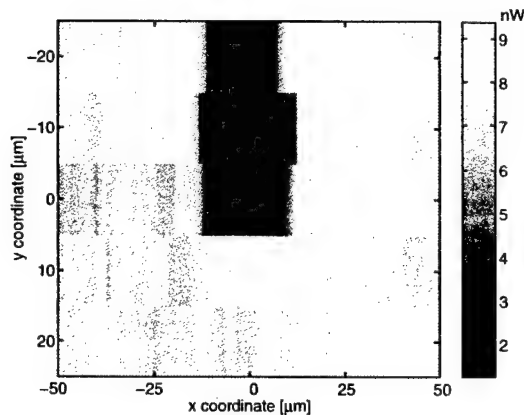

Fig. 4. Two-dimensional scanning confocal image (reconstructed intensity map) of the silicon microstructure obtained with the volume holographic microscope shown in Fig. 1

collector's spot size) and 1 mm. The intensity FWHM is (0.8 ± 0.1) μm for both curves, in close agreement with the value of ≈ 0.75 μm predicted by Eqs. (3) and (4). Note, however, that the pedestal of curve (b) is higher because of scattered light that is reaching the detector (i.e., the dynamic range of the measurement is slightly decreased). By contrast, the depth-discrimination capability of the confocal microscope [curves (c) and (d)] is degraded for the 1-mm aperture.

We used the pinhole-free confocal microscope to obtain a scanned image of the silicon microstructure, as shown in Fig. 4. The imaged portion contained a trench 20 μm wide and 5 μm deep. The reference surface for recording the hologram was outside the trench. The image of the trench corresponds to the dark region in Fig. 4, because the bottom of the trench is out of focus. We sampled only five planes along $\hat{y}$ and one along $\hat{z}$ to minimize inaccuracies that were due to the backlash of the translation stage and the decay of the hologram. A dense three-dimensional scan could have been obtained with a piezoelectric deflector and a fixed hologram.

In conclusion, we have demonstrated confocal scanning microscopy by use of a volume hologram as a shift-variant element matched to object depth. The dynamic range of volume holographic confocal imaging depends on the holographic diffraction efficiency (in our experiment it was $\approx 10^{-4}$) and is material limited. Single-hologram efficiencies as high as 100% have been demonstrated,[5] albeit with thinner materials and, hence, poorer Bragg selectivity. Volume holograms also permit the use of other imaging modes, e.g., color-selective (hyperspectral) tomographic imaging[6] or superresolution by use of complex filtering,[7,8] in combination with the pinhole-free confocal imaging principle.

References

1. T. Wilson and A. R. Carlini, Opt. Lett. **12,** 227 (1987); T. Wilson, in *Confocal Microscopy*, T. Wilson, ed. (Academic, San Diego, Calif., 1990), Chap. 3, pp. 93–141.
2. C. J. R. Sheppard and C. J. Cogswell, in *Confocal Microscopy*, T. Wilson, ed. (Academic, San Diego, Calif., 1990), Chap. 4, pp. 143–169.
3. N. P. Suh, A. C. Bell, and D. C. Gossard, Trans. ASME **100,** 127 (1978); N. P. Suh, *The Principles of Design* (Oxford University, New York, 1990).
4. C. Cohen-Tannoudji, B. Diu, and F. Laloë, *Quantum Mechanics* (Wiley-Interscience, Paris, 1977).
5. K. Meerholz, B. L. Volodin, B. S. Kippelen, and N. Peyghambarian, Nature **371,** 497 (1994).
6. G. Barbastathis and D. J. Brady, "Multidimensional tomographic imaging using volume holography," Proc. IEEE (to be published).
7. Z. S. Hegedus and V. Sarafis, J. Opt. Soc. Am. A **3,** 1892 (1986).
8. J. G. Walker, E. R. Pike, R. E. Davies, M. R. Young, G. J. Brakenhoff, and M. Bertero, J. Opt. Soc. Am. A **10,** 59 (1993).

Multidimensional Tomographic Imaging Using Volume Holography

GEORGE BARBASTATHIS, MEMBER, IEEE, AND DAVID J. BRADY, MEMBER, IEEE

Invited Paper

We propose the application of volume holography to four-dimensional (4-D) spatiospectral imaging. The proposed systems use materials and techniques developed for holographic data storage and interconnections to capture three-dimensional (3-D) spatial and one-dimensional (1-D) spectral information about a remote light source or scatterer. We analyze case studies of simple architectures using spherical-reference volume holograms as imaging elements in a fluorescence confocal microscope arrangement and demonstrate the equivalence of the holographic degeneracies with a slicing operation on the reconstructing incoherent source. We develop a general theoretical framework for the diffraction of random fields from volume holograms and show that the formulation can be used as an imaging design tool. Applications and future directions are also discussed.

Keywords—*Holography, microscopy, optical imaging, tomography.*

I. INTRODUCTION

The introduction of volume holography in a seminal paper by van Heerden [1] was soon followed by the discovery of appropriate materials through the effect of "optical damage" [2], which later became known as the photorefractive effect [3]. Since then, volume holograms have been popular in a number of subareas of optical information processing, namely data storage [1], [4]–[6], interconnects and artificial neural networks [7], [8], and communications [9]–[13]. To date, commercial applications of thick volume holograms are for spectral filtering [14] and three-dimensional (3-D) storage devices. In this paper we introduce a novel application of volume holography to multidimensional imaging.

Optical imaging is in the midst of a revolutionary shift from analog to digital systems. The most apparent aspects of this shift are the ubiquitous availability of digitized images and the use of digital networks to transmit images. Deeper aspects of the shift to digital techniques are only beginning to be explored, however. For example, the physical analogy between the detected field and the perceived object which is the basis of classical imaging need not be present in a digital system. In classical systems, a two-dimensional (2-D) focal plane pattern is used to represent the object in spite of the fact that the object is usually 3D. The goal in building a classical system is to make the field distribution on the sensor plane appear as similar as possible to the object viewed from the same perspective. Digital systems, in contrast, use sensor data to reconstruct the object in its native 3-D space. Since the digital system does not directly display sensor data, sensor data need not look like the object. The goal in designing a digital system is to maximize the detected object information so as to allow an accurate object model to be constructed. In many cases, it may not be possible to obtain simultaneously information on all object features. For example, capture of polarization data may preclude capture of spectral data or reduce spatial resolution, capture of temporal variations may limit 3-D resolution, etc. In view of these tradeoffs, digital systems are designed to optimize the capture of specific features of interest.

Imaging system design has been the primary subject of physical optics for millenia and the state of development of these systems is very high. While volume holograms can replicate the function of imaging system components, such as lenses, beam splitters, and spatial or spectral filters, holograms do not out perform conventional components for these functions. Volume holography as a tool is extremely attractive in emerging digital imaging systems, however, because volume holograms have more design degrees of freedom per unit system aperture than any other optical component. Design complexity allows volume holograms to extract more sophisticated features from fields, enabling sensor design to target features for object reconstruction.

Volume holograms for complex field transformation and feature extraction have been highly developed in the context of holographic storage and interconnection. Digital data storage, where each stored hologram corresponds to a page of information, is the most straightforward application. Despite this apparent simplicity, system geometry is extremely important to the capacity and function of data storage. On the next level of complexity, artificial neural networks have used holographic mappings for dendritic interconnections. Before the spectacular improvement of very large scale integration (VLSI) technology in the 1990's [15], [16], volume holograms were considered as a primary contender for the efficient storage and implementation of the massive interconnections needed for complex pattern recognition tasks. Many of the design considerations from data storage and neural net systems can be applied to the design of holograms for imaging applications. As will become evident in the remainder of the paper, some of the fundamental properties of holographic storage techniques, in particular the spatial selectivity and degeneracies of spherical reference volume holograms [17], can be applied verbatim to imaging. Even though the architectures we study here are different than the disk geometry of [17], the similarity simplifies our intuitive understanding of the problem.

The structure of the paper is as follows. Section II provides an extensive introduction to holographic storage and some of the issues arising when building a volume holographic system. In Section III we touch upon the primary issues arising in computational imaging systems and show as a simple example that the performance of the common fluorescence microscope arrangement improves when the collector lens is replaced by a volume hologram. In Section IV we analyze three simple implementations of a particular transformation, a matched filter to a point source, using volume holograms. We show that, when viewed on a flat camera detector, the diffracted field reconstructs a color-variant slice of the originating incoherent source, and we derive the slice shape as a function of the recording and reconstructing geometries. In Section V we develop a general procedure, formally equivalent to the Hopkins integral, for the calculation of diffraction of random optical fields from volume holograms. Our formulation leads to a design process, based on coherent mode decomposition, for constructing a volume hologram capable of shaping the coherence properties of the optical field arbitrarily, within the allowable degrees of freedom. We conclude in Section VI by discussing design considerations for multidimensional imaging systems, their markets, and applications.

II. HOLOGRAPHIC STORAGE

Holographic storage is motivated by high overall data capacity and parallel access. It was introduced by van Heerden [1], who first noted the similarity between X-ray diffraction from periodic crystal lattices and light diffraction from volume gratings and proposed utilizing this effect to superimpose and selectively retrieve multiple holograms in the same material volume, each hologram storing one page of information. The maximum number of resolvable voxels that can be stored inside a volume V at wavelength λ is V/λ^3. This corresponds to an order of 10 Tbits/cm^3 for green light. The parallelism, or the maximum number of resolvable pixels that can fit in a single page (i.e., an individual hologram), is bounded above roughly by $V^{2/3}/\lambda^2$. For green light, this is 0.4 Gbits/cm^2, with a data rate of several Gbits/s if the page size is actually 1 cm × 1 cm, and it takes no more than a few milliseconds to integrate each individual hologram on the detector. Neither of these upper bounds has ever been achieved in practice because of material and device limitations.

A typical holographic storage system is shown in Fig. 1. The hologram is recorded by illuminating a photosensitive material with the interference pattern formed by two coherent light beams, the reference and signal. The signal beam contains the information to be stored in the form of transverse phase or amplitude modulation of the beam profile, imposed by a spatial light modulator (SLM). The reference beam contains no information, except the "identity" of the hologram. For example, in the most common form of holographic storage, called "angle-multiplexing" [18]–[20], which is depicted in Fig. 1, the reference beam for the mth hologram is a plane wave incident at angle θ_m. After the exposure is complete, each hologram ideally contributes an equal amount of spatial modulation to the refractive index of the material. The mth hologram is then accessed selectively by illuminating the exposed material with the corresponding plane wave at angle θ_m. If the original recording reference beams were appropriately spaced, then the diffracted light contains significant reconstruction from the mth hologram only. The remaining holograms are Bragg mismatched, i.e., they are read out by the incident beam, but their reconstructions, when integrated over the entire volume of the material, cancel out to zero. In the common configuration of Fig. 1, the angular separation between adjacent holograms must be equal, approximately, to an integral multiple of

$$\Delta\theta = \frac{\lambda}{2L\sin\theta}. \tag{1}$$

This quantity is known as angle Bragg selectivity. Since $\Delta\theta$ is proportional to λ/L, the selectivity improves by using shorter wavelengths or thicker materials. It is important to note that the multiplexed holograms share the entire volume of the recording material; therefore, holographic storage is fundamentally different than layered volume storage methods, such as the digital video disk (DVD) and two-photon storage [21]. One might think of the process of Bragg matching a single hologram in the presence of multiple holograms sharing the medium as similar to tuning a receiver to a radio station; the matching angle θ_m corresponds to the resonance frequency of the receiver, and the Bragg separation $\Delta\theta$ corresponds to the quality factor Q that determines the receiver bandwidth.

Angle multiplexing has been by far the most popular technique in experimental demonstrations. The angular
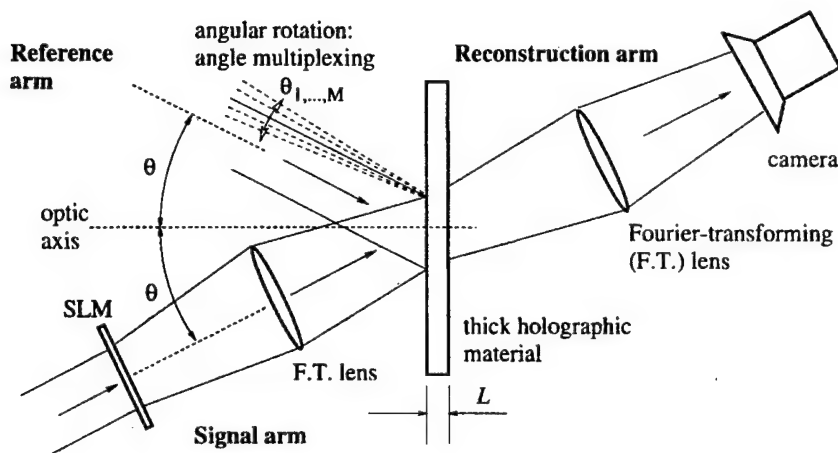
Fig. 1. A common holographic memory architecture. The mth hologram is recorded by the interference of the beams from the "reference" (incident at θ_m) and "signal" arms (incident at angle θ). The information to be stored is modulated on the signal wavefront by the SLM. When the hologram is illuminated by a plane wave incident at angle θ_m from the reference arm, the mth stored page is diffracted into the "reconstruction" arm and is focused onto the camera. This scheme is called "angle multiplexing."

deflection required to record and access different holograms has been implemented by electromechanical actuation of off-the-shelf commercial mirrors [5], [6], [22]–[24], magnetic actuation of micromachined mirror flaps [25], acoustooptic deflection [26], [27], and liquid-crystal-based electrooptic deflection [28], [29].

Alternative multiplexing methods have also been devised. For example, using a plane wave reference beam, the reconstruction is also sensitive to the wavelength of the incident beam [18], [19], [30], [31]. With widely tunable visible and near-IR lasers becoming more common, compact, long lived, and affordable, the elimination of the need for an angular deflector makes wavelength multiplexing a more attractive choice. Another method without mechanical addressing requirement is phase-code multiplexing [32] where the reference beams are implemented as a set of orthogonal codes and addressed using a phase SLM. Even simpler is the implementation of shift multiplexing, which requires a reference beam that is either a collection of plane waves with a regular relative angle displacement [33] or which is a spherical wave [17]. In both cases, individual holograms are accessed by relative translation between the reference and the recording medium. The required shift between adjacent holograms is typically of the order of a few micrometers. The shift could be implemented acoustooptically, but mechanical translation is simpler and well characterized because of the popularity of optical storage disks [34]. Therefore, the latter has been the method of choice in high-capacity experiments [35], [36]. The properties of spherical reference volume holograms will be revisited in detail in Section IV.

The techniques mentioned so far make use of Bragg mismatch to multiplex holograms. Further increase in capacity may be obtained by synthetically increasing the aperture of the hologram using the motion of the reference beam. In the holographic storage jargon, these techniques are referred to as "fractal" [37]. Recent implementations include the

"peristrophic" multiplexing method [38] and the hybrid angle-wavelength multiplexing method [39]. The aperture increase is effected by use of the degeneracy effects that will be derived for some specific geometries, but in quite a different context, in Section IV-C.

With a wide choice of well-understood multiplexing techniques available, the next critical system issue is the material [40], which is determined by the application. We consider photorefractive and photopolymer materials only, because so far they have been the most popular in experiments of erasable and permanent holographic storage, respectively. A complete review of available holographic storage materials [41] is outside the scope of this paper.

Photorefractive crystals, such as Fe-doped $LiNbO_3$, $Sr_xBa_{1-x}NbO_3$ (SBN:x), and $BaTiO_3$ were the first materials to be used for holographic storage [3]. During recording, the refractive index change occurs via the electrooptic effect after a spatially varying space-charge field is established in the crystal from the diffusion or drift of photo-excited charges away from the illuminated regions [42]–[44]. The space-charge field sustains itself after removal of the recording beams but decays because of thermal electronic excitation in the dark, or uniform photo-excitation during hologram readout. Decay occurs also as a result of superimposing more holograms in the same location of the material. As a result of the erasure of existing holograms when new holograms are recorded, the dynamic range of the material is not fully utilized, and the diffraction efficiency (defined as the portion of the reference beam power diffracted into the hologram) of $M \gg 1$ equal-strength holograms is [45], [46]

$$\eta(M) = \frac{(M/\#)^2}{M^2}. \qquad (2)$$

The parameter $M/\#$ (pronounced "M-number") depends highly on material parameters, such as absorption coefficient, doping levels, recombination lifetimes, etc., but

also on the beam profiles and intensities and stability of the experimental arrangement; it is, therefore, a system parameter [46]. Typical photorefractives have $M/1$ or less, but there are exceptions [22].

Photorefractive holograms are semipermanent, and, therefore, appropriate for optically erasable, rewritable, and refreshable random access memory architectures [5], [22]–[24], [29], [47], or when dynamic holography is required, e.g., two-wave mixing [48], [49], phase conjugation [50], [51], optical novelty filters [52], self-waveguiding [53], [54], etc. Photorefractives are often used also in applications that require permanent storage, because the crystal thickness can be large (several millimeters or centimeters), thus providing high capacity. A number of techniques exist for recording permanent holograms in photorefractives and include thermal fixing [55]–[60], electrical fixing [61]–[65], two-lambda readout [66]–[70], and two-photon recording [71]–[75]. A comprehensive review of nonvolatile photorefractive storage is given in [76].

A different class of holographic recording mechanisms is based on photochemical changes initiated by exposure to the recording beams. The most common example is photoinduced polymerization in the DuPont polymer HRF-150 [77]–[79], where recording occurs as refractive index modulation because of density changes in the exposed areas; it is permanent and does not significantly degrade over time. Despite the different recording mechanism, the diffraction efficiency as a function of number of superimposed holograms still follows the rule (2). The HRF-150 has been demonstrated to have approximately $M/6$, and has been used successfully in a number of high-capacity demonstrations of holographic storage [35], [80], [81].

The selection of material and multiplexing technique depends on the application. Storage in photopolymers is permanent, hence they target read-only (ROM) or write-once-read-many (WORM) storage applications. Unfortunately, the thickness of photopolymer films is limited by considerations of mechanical stability and optical quality. The highest capacity ever achieved in the DuPont polymer is 12 bits/μm^2 [35] using shift multiplexing with a 100 μm thick film. This surface density is higher than the DVD-ROM by a factor of two. Recently, samples of thickness up to 5 mm were fabricated using a poly(methyl-methacrylate) (PMMA) polymer matrix to host the photosensitive material phenanthrenequinone (PQ) [82]–[84]. Theoretical calculations [17], [85] show that the achievable density at 5 mm hologram thickness is as high as 200 bits/μm^2. Therefore, PQ-doped PMMA seems promising as a replacement to the DuPont HRF-150 polymer and nonvolatile photorefractive storage for permanent high-density holographic memories.

Other systems issues that are important for holographic storage are page-oriented error correction [86]–[88] and channel modulation [89], [90], pixel matching [91] (i.e., minimizing aberration distortion by using unit magnification in the optical system between the SLM and the detector), and the location of the hologram with respect to the imaging system [92] (i.e., whether the hologram should
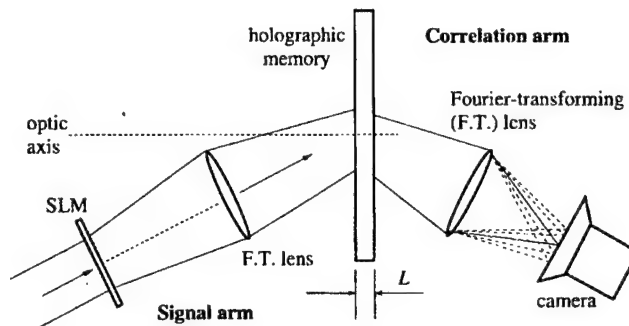
Fig. 2. Operation of the angle-multiplexed holographic memory of Fig. 1 in correlator mode.

be located on the focal or pupil plane of the imaging system that maps the SLM on the detector plane). A complete review of these issues is outside the scope of the present paper.

The function of volume holograms as correlators [93] has been traditionally an important application of holographic memories oriented toward optical pattern recognition [94]–[96]. Suppose M patterns f_m $(m = 1, \ldots, M)$ are stored in a holographic memory. If the memory is illuminated by a new pattern g along the path of the signal beam, and a Fourier-transforming lens is placed on the continuation reference path (see Fig. 2), then at the focal plane one obtains the correlations $g \star f_m$ of the novel pattern with all the stored patterns at once. The parallel correlation operation is obtained at the expense of losing shift invariance in one dimension at the output plane. This mode of operation of a holographic memory has been successfully used in a number of applications [97]–[99]. In Sections IV and V we will show that the volume hologram correlates its internal modes with the input field. This function is useful as an imaging operation.

III. 3-D Imaging and Volume Holography

A. Types of Imaging Systems

An optical imaging system transfers information about an object to the user, using light as information carrier. The amount and quality of the transmitted information is determined by the propagation properties of light. Free space propagation has the effect of delocalizing the object features, "blurring" the image. Optical elements, such as lenses, are used to compensate propagation and recover the object features locally or bring the image "in focus."

Most imaging instruments assume planar objects, i.e., objects that can be described by a two-variable function defined on a surface transverse to the optical axis (see Fig. 3). The optical system performs an analog linear transformation on the transverse field intensity distribution, and the image appears at the final detection stage. The imaging task is more demanding for 3-D objects, because it requires compensation of light propagation effects in three dimensions. Unfortunately optical instruments are geared to handle planar rather than volumetric objects, and optical detectors are also typically planar (2D). Three-

(a)

(b)

Fig. 3. (a) A planar imaging geometry. (b) A volume imaging geometry. The dashed lens indicates that the imaging system is usually more complicated than a single lens.

imaged point
(target)

Fig. 4. Principle of the confocal microscope arrangement. In experimental configurations, the sample is sometimes: 1) reflective, when the system is folded, sharing the same lens as objective and collector, or 2) fluorescent, when the radiation emitted by the sample is at a longer wavelength and the illuminating radiation is blocked by a color filter.
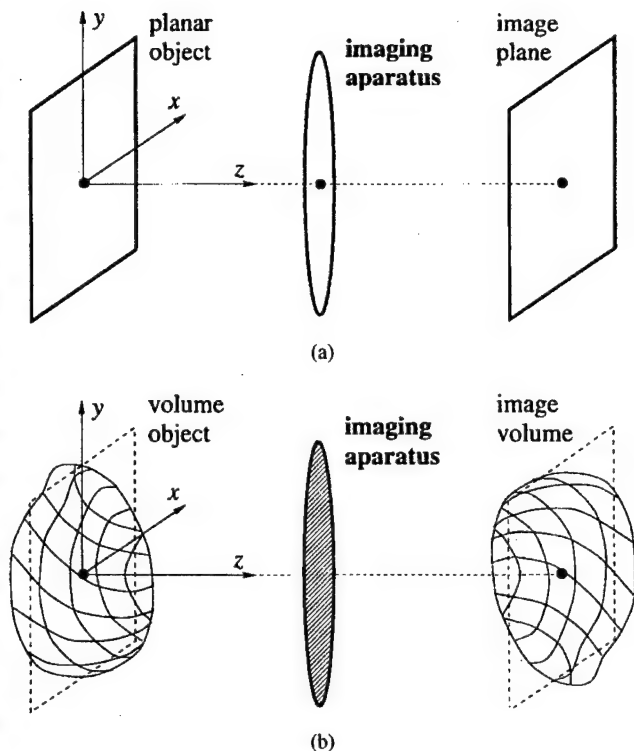
dimensional imaging requires the acquisition of sets of lower-dimensional intensity measurements (2-D or point measurements) and the subsequent formation of the image from these measurements. One selection for the intermediate measurements may be 2-D images of slices of the 3-D object. The imaging system is then called "tomographic." The content of the intermediate measurements, however, may not resemble the object at all. Then, a more complicated transformation is required to recover the image. This class of "computational imaging" systems is quickly becoming more popular as the available digital processing power increases.

Three-dimensional optical imaging schemes may be classified into five broad categories: scanned systems; scene analysis systems; projective systems; interferometric systems; and modal systems. Scanned systems include laser spot scanners, confocal microscopes, and laser fluorescence microscopes. These systems are effective but slow, since the volume data are acquired one spot at a time. Scene analysis systems combine expert systems and geometry to computationally reconstruct objects; they require substantial prior object knowledge. Projective systems combine ray optics and inverse-Radon or similar transforms to reconstruct objects and work best with high depth of field optical components. Interferometric systems include holographic schemes and coherence tomography; they are very powerful and general but are subject to noise concerns. Modal systems take the most general approach, detecting the state of all optical modes and attempting a computational inversion. All five imaging system classes require new approaches to

optical design and benefit from spatial and spectral filtering. Volume holographic elements can substantially improve any type of imaging system. In this paper we will illustrate this potential for the cases of confocal microscopy, tomography, and coherence imaging. The following paragraphs describe the traditional approaches to confocal microscopy and coherence imaging in more detail.

The confocal microscope, invented by Minsky [100], operates by the lowest dimensional measurements possible, i.e., point measurements. A confocal microscope is sketched schematically in Fig. 4. It constructs a 3-D image by scanning the volume of the specimen and obtaining the emitted intensity values one point at a time. The geometry of the optical system is such that light emitted locally from a very small portion of the object only is allowed to reach the detector. The rest of the light is rejected by the aperture at the detector pupil. The proportional light contribution to a single measurement as function of object coordinates is equivalent to the 3-D point-spread function (PSF) of the system; it can be calculated accurately under various aberration conditions using Fourier optics [101], [102]. Confocal microscopy has been implemented in many different variants for improved light efficiency or resolution, e.g., differential interference [103], fluorescence [104], two-photon [105], etc.; it has been spectacularly successful, primarily in various applications of biological and biomedical imaging.

Coherence imaging (Fig. 5) is an example of computational imaging that relies on global, rather than local, measurements. It is based on a fundamental result, derived independently by van Cittert and Zernicke [106], [107], which states that the degree of statistical correlation of the optical field in the far zone, expressed as a complex function over the exit pupil of the imaging system, is the Fourier transform of the object intensity distribution. Therefore, the object can be recovered by measuring the coherence function through interferometry and then inverse-Fourier transforming the result. The application of the van Cittert–Zernicke theorem in the radio frequency spectral

Fig. 5. A system for implementing coherence imaging with a "rotational shear interferometer" (after [116]). The roof prisms rotate the image antisymmetrically about the two axes, and the two versions are interfered at the camera plane.

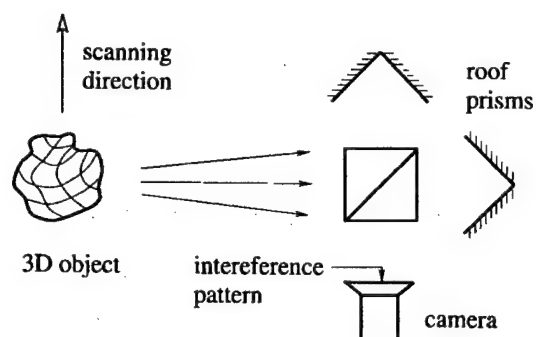region is the basis of radio astronomy [108], which yields by far the most accurate images of the most remote cosmic objects. The most common formulation of the theorem relates the mutual coherence in a plane at infinity to a 2-D source intensity distribution, but extentions to 3-D sources have been derived by various authors [109]–[112]. The far-field version of the extended van Cittert–Zernicke theorem was recently implemented experimentally [113]–[115]. A full generalization of the theorem has also been developed and experimentally implemented to allow Fresnel zone reconstruction in projective coordinates [116].

In either of these systems the detected intensity is shaped by the response of the imaging elements (in the confocal microscope, the intensity measurement results from the field received at a single point, whereas in coherence imaging the measured intensity is the result of interference between two or more optical paths). A volume hologram is a more general design tool. One may think of it as an element that modifies an optical beam continuously along an entire volume. As such, it can be designed to perform spatial filtering operations similar to the confocal microscope, but it is more sophisticated because of the additional third degree of freedom, as we show in an example in Section III-C. An even more extended operation available by the volume hologram is a spatiospectral mapping between points in the object and points on the detector. As we will see in Section IV-C, the so-called "degeneracy" properties [37] of the volume hologram provide this mapping; the recording geometry is the tool that allows the designer to shape the map structure. The most general usage of volume holograms for imaging is by way of mixing the modes of the field generated by the object with the modes of the hologram through the effect of volume diffraction. Whereas the previous examples can be classified as subcategories of modal imaging, volume holography allows arbitrary shaping the coherence properties of the scattered field. The formal development of this design technique is given in Section V.

B. Imaging System Design

The impact of design choices in individual optical components on system performance is a critical issue for 3-D

imaging system design. Lens behavior, for example, has been well characterized in a large variety of imaging conditions, and lens design is an art in itself. Confocal microscopy, along with a large number of high-performance imaging techniques, make good use of advances in lens design. On the other hand, in lensless imaging systems (such as coherence imaging, mentioned above), one tries to get away from the complexity of lens design by using simpler elements (mirrors, prisms) to form interference patterns, and subsequently one uses the computational power of digital computers to apply transformations (Fourier transforms, Fresnel transforms, and possibly nonlinearities) on the detected image intensity in order to recover the 3-D object. With the exception of digital computations, the design of all other imaging system elements is constrained by machining accuracy limitations. Digital transformations themselves are limited by the requirement of reasonable computation time. Therefore, part of the imaging design problem is to achieve a successful balance in splitting the imaging transformations to analog ones, performed by the optical elements, and digital ones, performed by computers, according to the individual capabilities of each component.

A class of optical elements that allow considerable flexibility in their optical response is, of course, holograms. A hologram is determined either by the profiles of the two optical beams that interfere to record it or can be fabricated by etching a waveform on a suitable material (typically glass). In either case, several sophisticated devices are available for determining the hologram response. For example, in the former case, SLM technology allows spatial amplitude and phase modulation of optical beams to resolution down to 10 μm; in the latter, photolithography and electron beam patterning have been used to generate very sophisticated diffractive optical elements for communications, display, and other applications. Holograms have not been very popular as optical elements in practical imaging systems. A notable exception is holographic interferometry [117] and two-wavelength interferometry [118], [119], where the hologram does not function as a fixed imaging element, but rather as a sophisticated detector that captures phase properties of the object. Bertero and collaborators [120]–[126] have proposed a method of superresolving confocal microscopy using diffractive elements calculated based on singular system theory.

We propose to use volume holograms as optical imaging elements for one main reason: a volume element provides a larger number of degrees of freedom in defining the optical response, compared to a surface element (e.g., a thin hologram) of the same aperture. This is intuitively obvious from dimensional arguments and was proven formally in [127] and [128] using the modal properties of electromagnetic fields. We will not repeat the formal arguments here but point out the desirable and undesirable features of volume holography that should be taken into account in the design process. The main price to pay for the advanced design flexibility is that the control problem of defining the hologram response (i.e., "programming" the volume hologram) becomes considerably more difficult and

is accomplished at the expense of diffraction efficiency [128]. Other considerations that follow from the description of Section II are:

1) volume holography provides enormous storage capacity; therefore, a large number of degrees of freedom is available to the designer for shaping the optical response and improving the quality of the image;

2) the capacity goal should be achieved by using as small a number of holograms as possible in order to maintain high individual diffraction efficiency for each hologram;

3) the recording of volume holograms is an expensive, material-limited process that should not be performed in real time; it is better, therefore, to use volume holograms as fixed elements that have been predesigned, fabricated in the factory, and delivered to the user, rather than as dynamic elements modifiable in real time.

What function should the volume hologram perform inside an optical imaging system? Unlike other optical elements, the range of possible responses by volume holograms allows them to perform several functions. We conclude this section by giving an example of a volume hologram as part of a confocal imaging system. The more complicated nature of the hologram's response is fully developed in Section IV-C for several recording geometries. In these cases, the volume hologram acts as a local imaging system by isolating specific light contributions arising from spatial and spectral bands of the object and mapping them onto a 2-D detector. At the end of Section V we will see that a volume hologram may also be designed to act as a global imaging instrument that forms correlations between the light modes emitted (or scattered) by the object and the modes of the hologram.

C. Example: Confocal Imaging with a Volume Holographic Collector

Consider again the confocal imaging system of Fig. 4. The most common performance measure of such a system is "resolution"; i.e., the size of the minimum resolvable element within the object volume. This is equivalent to the volume where the 3-D-PSF of the confocal imaging takes significant values. Ideally, the 3-D PSF is a δ function and the resolution is infinite, but in real-life systems it is nonzero over a finite volume. The confocal arrangement achieves a tight 3-D PSF by: 1) illuminating the point of interest inside the object (the target) by a tightly focused beam, produced by the objective lens, and 2) re-imaging through the collector lens the radiation from the target onto a small pinhole aperture in front of the detector. Thus, point radiators other than the target are doubly inhibited: 1) they are illuminated by an extended low-intensity beam, whereas the target is illuminated by the high-intensity beam waist and 2) the radiation they produce is rejected by the 3-D PSF of the collector because they are away from the focal point.
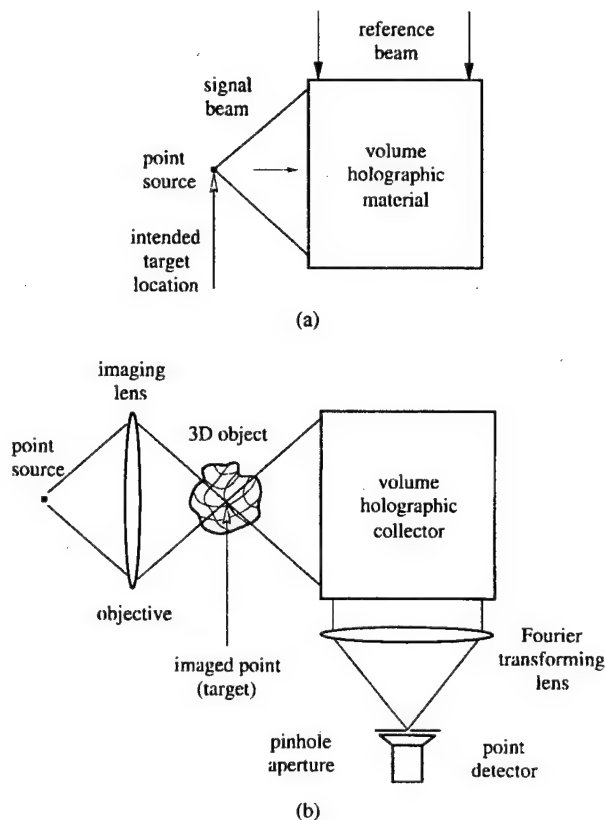
Fig. 6. Confocal imaging arrangement with the collector lens replaced by a volume hologram and a Fourier transforming lens: (a) geometry for recording the volume hologram and (b) confocal imaging arrangement.

Consider now the modified confocal imaging system of Fig. 6(b), where the collector has been replaced by a volume hologram and a Fourier-transforming lens. The volume hologram has been recorded by the interference of a spherical wave originating from the intended target location, and a plane wave oriented normally with respect to the optic axis of the spherical wave, as shown in Fig. 6(a). This recording arrangement is known as "90° geometry," and has been popular in a number of holographic storage demonstrations [5], [6], [22]–[24]. In our imaging configuration [Fig. 6(b)], the volume hologram captures the radiation emitted by the object after illumination by the focused input beam produced by the objective. The diffracted light propagates in the direction shown in Fig. 6(b) and is then captured and Fourier transformed by the lens. The pinhole-sized detector is placed at the focal point of the lens, i.e., it captures the dc component of the diffracted field.

Formally, the volume hologram+lens arrangement forms the correlation between the field emitted by the object and the original signal beam [the spherical wave of Fig. 6(a)]. Radiation emitted from the target position at the recording wavelength is identical to the recording signal and is reconstructed on the detector. Radiation emitted from different positions and at different wavelengths (if the object happens to be polychromatic) does not correlate well with the recording signal and is not reconstructed. The calculation of the diffracted field as a function of the reconstructing object
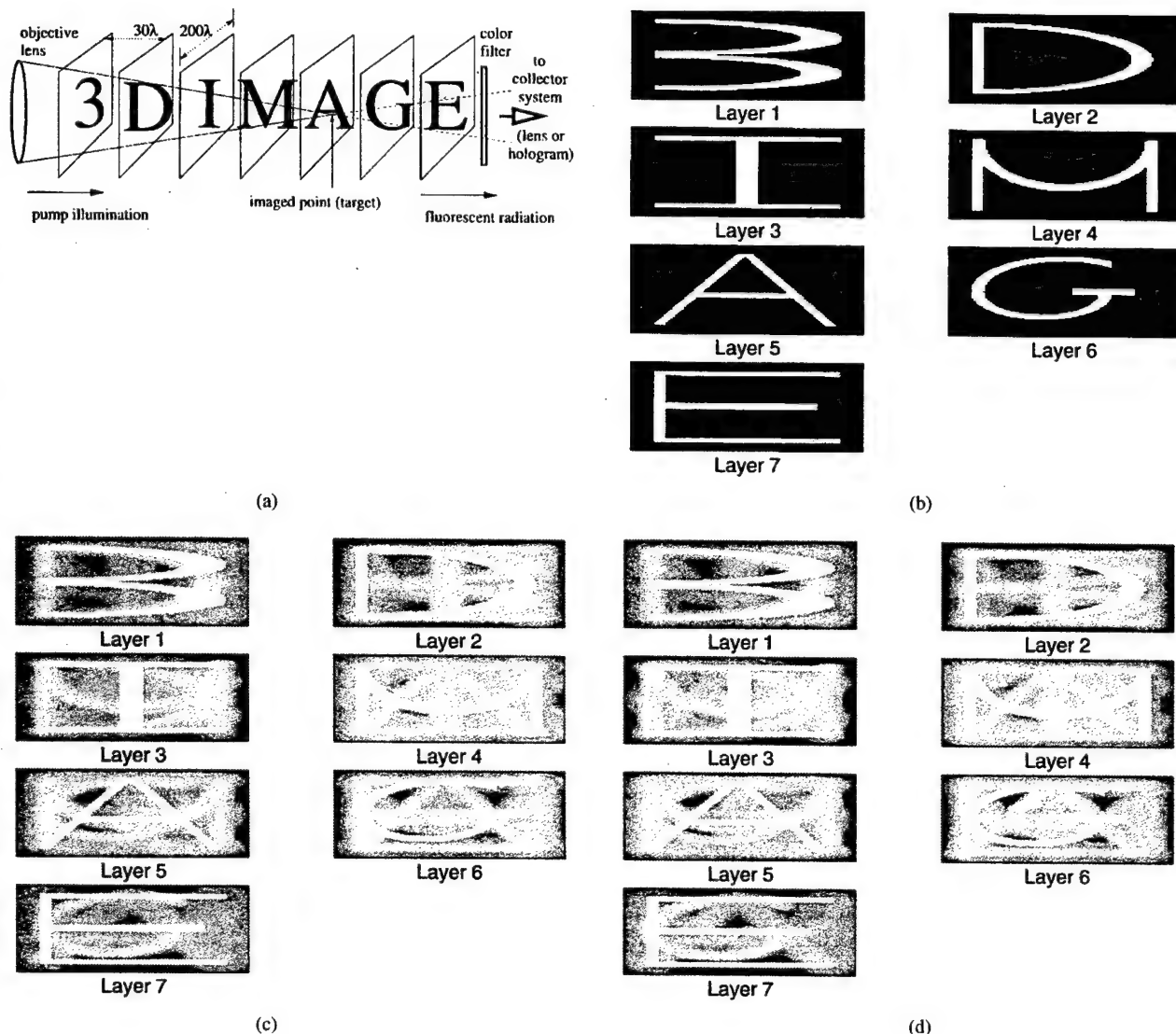
Fig. 7. Numerical comparison of fluorescence confocal imaging with a lens collector and volume holographic collector. For simplicity, the fluorescence wavelength is taken to be equal to the probe wavelength. (a) Geometry used in the simulation. Each plane is modeled as a 100×100 grid of incoherent radiators at the same wavelength λ. (b) Shape of the original object. (c) Fluorescence confocal reconstruction using a lens collector of aperture radius 1500λ. located 2500λ from the plane of the letter "M." (d) Fluorescence confocal reconstruction using a volume holographic collector with aperture radius 1500λ, thickness 3000λ, with its center located 2500λ from the plane of the letter "M."

for this geometry is given in detail in Section IV-B. There we will find out that some parts of the uncorrelated radiation actually are reconstructed, but not at the focal point of the lens. Detecting these reconstructions in an organized way allows the performing of interesting slicing operations on the object. These will be explained in Section IV-C. For the purposes of this section, it suffices to note that, in isolating the radiation emanating from the target point and rejecting the rest, the volume holographic collector is more efficient than an equivalent lens collector. The reason is understood immediately upon comparing the 3-D PSF's of the confocal arrangement with a lens collector as opposed to a volume hologram. The 3-D-PSF calculation is performed using volume diffraction theory in Section IV-B. There, it is shown that the width of the main lobe is the same in both

cases; however, the sidelobes are significantly suppressed in the case of the volume holographic collector.

A numerical example demonstrating the importance of the side-lobes is given in Fig. 7. In this example, we numerically reconstructed a fluorescent 3-D incoherent object with the two cases of confocal microscope with a regular lens and volume holographic collector. The resolution was close to the borderline resolution allowed by the numerical aperture of the lens collector. From the reconstructions we see that the lens collector accumulates noise from power diffracted by the sidelobes; this is absent from the volume holographic reconstruction. The signal-to-noise ratios, computed as the quadratic error between object and image normalized to the total image intensity, were ≈ 1500 and ≈ 3500, respectively, for the lens and volume

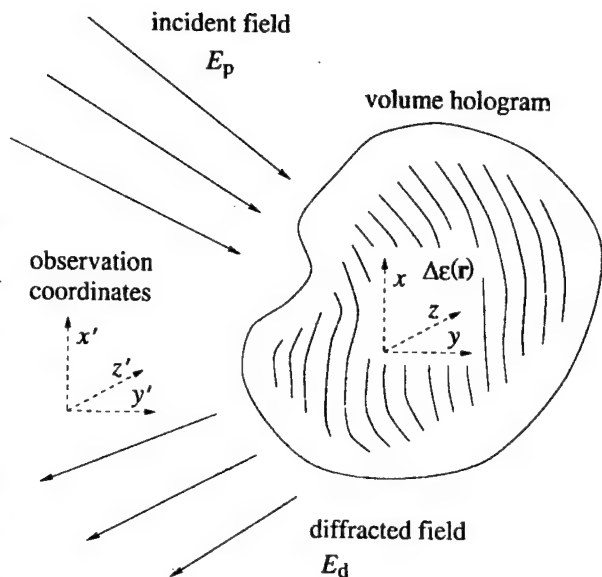**Fig. 8.** Schematic of a generic volume diffraction geometry. The probe field $E_p(\mathbf{r})$ and index modulation $\Delta\varepsilon(\mathbf{r})$ are expressed in the xyz coordinate system. For notational clarity, we use a different $x'y'z'$ coordinate system for the diffracted field $E_d(\mathbf{r})$.

Fig. 9. Simplified holographic recording geometries considered in this paper: (a) reflection geometry and (b) 90° geometry. The reference and probe beams are always spherical waves. We cover both cases when the signal is either spherical wave or a plane wave for (a) in Sections IV-A and IV-C and the case of plane wave signal only for (b) in Sections IV-B and IV-C. The reconstructed beam depends on the relative position and wavelength of the reference and probe beams, the nature of the signal beam, and the shape of the hologram.

holographic collector. This improvement is not dramatic, but the design can be combined with matched filtering similar to Bertero's decomposition method [120]–[126], yielding even better results. This last step is not described in this paper.

IV. DIFFRACTION PROPERTIES OF VOLUME HOLOGRAMS WITH SPHERICAL REFERENCE

As we discussed in Section II, in most holographic materials the recording of a volume hologram is accomplished through modulation of the refractive index. This is expressed as a function $\Delta\varepsilon(\mathbf{r})$ of the space coordinate $\mathbf{r}$, for $\mathbf{r} \in V_{\mathcal{H}}$, where $V_{\mathcal{H}}$ is the volume occupied by the holographic material. A generalized version of the volume diffraction geometry, valid for all the calculations of this section, is given in Fig. 8. When the hologram is illuminated by a probe field $E_p(\mathbf{r})$, the diffracted field $E_d(\mathbf{r}')$ is found as the solution to Maxwell's equations in an inhomogeneous medium, with refractive index as given above. The solution is simplified if we assume that the magnitude of the modulation is much smaller than the unmodulated refractive index ε_0

$$|\Delta\varepsilon(\mathbf{r})| \ll \varepsilon_0, \qquad \mathbf{r} \in V_{\mathcal{H}} \tag{3}$$

because the weak diffraction approximation (also known as "Born's approximation") can then be applied. The diffracted field is given by

$$E_d(\mathbf{r}') = \iiint\limits_{V_{\mathcal{H}}} E_p(\mathbf{r})\Delta\varepsilon(\mathbf{r}) \times \frac{\exp\{ik|\mathbf{r}-\mathbf{r}'|\}}{|\mathbf{r}-\mathbf{r}'|} \, d^3\mathbf{r} \tag{4}$$

where $k = 2\pi/\lambda$ is the wavenumber and the last term in the integrand is recognized as the scalar Green's function for free space. The derivation of (4) from Maxwell's equations is beyond the scope of this paper.
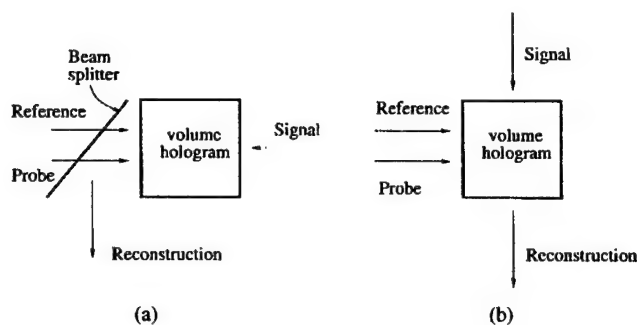
Equation (4) has a simple interpretation. Assume that the volume grating is composed of infinitesimal scatterers, the strength of the scatterer located at $\mathbf{r} \in V_{\mathcal{H}}$ being $\Delta\varepsilon(\mathbf{r})$. Then the diffracted field is the coherent summation of the fields emitted by all the scatterers when they are excited by the incident field $E_p(\mathbf{r})$. Naturally, this picture omits higher order scattering, i.e., fields generated when the field scattered from one infinitesimal scatterer reaches other infinitesimal scatterers. This omission, though, is consistent with the weak scattering approximation, which says that these higher order effects are even weaker and, therefore, negligible.

Expression (4) is computationally efficient when spherical waves are involved in the recording of the hologram, as we will see in the next two sections. For other types of fields, a representation of the diffracted field and the grating in wave-vector space works better but is beyond the scope of this paper. For a more complete treatment, the reader is referred to [129].

We will be examining two volume holographic geometries, shown in Fig. 9. In the "reflection geometry" [Fig. 9(a)] the reference and signal beams are incident on two opposite faces of the holographic material and (approximately) counterpropagating. Upon reconstruction, the probe beam is incident in the direction of the reference and the diffracted beam is generated as extension of the signal, i.e., it is counterpropagating, on the same side of the medium as the probe beam. A beam splitter is used to separate the reconstruction from the probe. In the "90° geometry" [Fig. 9(b)] the reference and signal beam are incident on two normal faces of a cube-like recording medium. Again, the probe is incident from the same direction as the reference, and the reconstruction appears as a continuation of the signal, but no beam splitter is required in this geometry.

In the next two sections, we derive the basic formulas that give the diffracted field as function of the output coordinates $\mathbf{r}'$, the recording beams and the geometry of the hologram, for the reflection and 90° geometry, respectively.
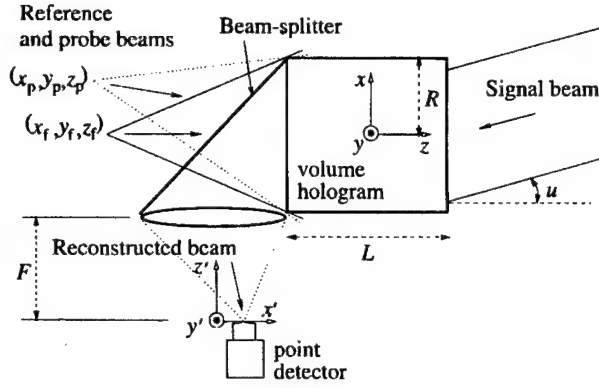
Fig. 10. Schematic of the reflection geometry with spherical wave reference and plane wave signal beams.

In Section IV-C we solve for the locus of probe points and wavelengths that generate maximum reconstructed intensity in the (arbitrarily defined) output plane. The resulting construction is called the "degeneracy surface" of the volume hologram and is important because it specifies the portion of the object that is "visible" by the hologram for imaging purposes.

A. Reflection Holograms

First we consider the geometry of Fig. 10, with a *plane wave* signal beam. The reference beam used for recording is a spherical wave at wavelength λ_f produced by a point source at $\mathbf{r}_f = x_f \hat{\mathbf{x}} + y_f \hat{\mathbf{y}} + z_f \hat{\mathbf{z}}$. We express this wave in the paraxial approximation, as

$$E_f(\mathbf{r}) = \exp\left\{ i2\pi \frac{z - z_f}{\lambda_f} + i\pi \frac{(x - x_f)^2 + (y - y_f)^2}{\lambda_f(z - z_f)} \right\}. \tag{5}$$

Note that we have neglected a term of the form $1/\lambda(z - z_f)$ because it varies with z much slower than the exponential term. Such slowly varying terms will be neglected from here on. The signal beam is a plane wave propagating at angle $u \ll 1$ with respect to the $\hat{\mathbf{z}}$ axis. In the paraxial approximation, it is expressed as

$$E_s(\mathbf{r}) = \exp\left\{ -i2\pi \left(1 - \frac{u^2}{2} \right) \frac{z}{\lambda} + i2\pi u \frac{x}{\lambda} \right\}. \tag{6}$$

The modulation of the material refractive index resulting from exposure to beams E_f, E_s given by

$$\Delta\varepsilon(\mathbf{r}) = |E_f(\mathbf{r}) + E_s(\mathbf{r})|^2. \tag{7}$$

Out of the four terms in the interference pattern, we will insert only $E_f^*(\mathbf{r})E_s(\mathbf{r})$ in the volume diffraction equation (4) for the remainder of this section. The remaining three terms are Bragg mismatched and do not diffract significantly.

The probe field is a spherical wave at wavelength λ_p emanating at $\mathbf{r}_p = x_p \hat{\mathbf{x}} + y_p \hat{\mathbf{y}} + z_p \hat{\mathbf{z}}$. The expression for the probe field is

$$E_p(\mathbf{r}) = \exp\left\{ i2\pi \frac{z - z_p}{\lambda_p} + i\pi \frac{(x - x_p)^2 + (y - y_p)^2}{\lambda_p(z - z_p)} \right\}. \tag{8}$$

To find the diffracted field at the detector coordinates $\mathbf{r}'$ (located near the focus $\mathbf{r}_s$ of the signal beam) we will use Born's diffraction formula (4). We simplify by assuming that the holographic medium is disk shaped with radius R in the xy plane, and thickness L along the z direction, and making the paraxial approximation, i.e., assume that R is smaller than any longitudinal distance that the fields propagate. We then obtain

$$E_d(\mathbf{r}'') = \iiint E_p(\mathbf{r})\Delta\varepsilon(\mathbf{r}) \operatorname{circ}\left(\frac{\sqrt{x^2 + y^2}}{R} \right) \operatorname{rect}\left(\frac{z}{L} \right)$$
$$\cdot \exp\left\{ i2\pi \frac{z'' - z}{\lambda_p} + i\pi \frac{(x'' - x)^2 + (y'' - y)^2}{\lambda_p(z'' - z)} \right\} d^2\mathbf{r}. \tag{9}$$

The field reaching the detector is obtained after a Fourier-transforming operation applied by the lens on $E_d(\mathbf{r}'')$, i.e.,

$$\tilde{E}_d(\mathbf{r}') = \iint_{-\infty}^{+\infty} E_d(\mathbf{r}'')$$
$$\cdot \exp\left\{ -i2\pi \frac{x'x'' + y'y''}{\lambda_p F} \right\} dx'' \, dy'' \tag{10}$$

where F is the focal length of the lens and constant phase factors have been omitted. The limits of integration in (10) are taken to be infinite by assuming that the aperture of the Fourier transforming lens is larger than the effective aperture imposed on the diffracted field by the transverse size R of the volume hologram. In other words, we assume that the volume hologram defines the aperture of the system. Under this condition, we can substitute (9) into (10) and perform the x'', y'' integrations right away, obtaining

$$\tilde{E}_d(\mathbf{r}') = \iiint \exp\left\{ i\pi A(z)(x^2 + y^2) \right\}$$
$$\cdot \exp\left\{ -i2\pi[B_x(z)x + B_y(z)y] \right\}$$
$$\cdot \exp\left\{ i\pi C(z) \right\} \operatorname{circ}\left(\frac{\sqrt{x^2 + y^2}}{R} \right)$$
$$\cdot \operatorname{rect}\left(\frac{z}{L} \right) d^2\mathbf{r} \tag{11}$$

where the coefficients $A(z)$, $B_x(z)$, $B_y(z)$, $C(z)$ are given by

$$A(z) = \frac{1}{\lambda_p(z - z_p)} - \frac{1}{\lambda_f(z - z_f)} \tag{12}$$

$$B_x(z) = -\frac{x_p}{\lambda_p(z - z_p)} + \frac{x_f}{\lambda_f(z - z_f)} - \frac{x'}{\lambda_p F} + \frac{u}{\lambda_f} \tag{13}$$

$$B_y(z) = -\frac{y_p}{\lambda_p(z - z_p)} + \frac{y_f}{\lambda_f(z - z_f)} - \frac{y'}{\lambda_p F} \tag{14}$$

$$C(z) = \frac{x_p^2 + y_p^2}{\lambda_f(z - z_p)} - \frac{x_f^2 + y_f^2}{\lambda_f(z - z_f)}$$
$$+ \left[4\left(\frac{1}{\lambda_p} - \frac{1}{\lambda_f} \right) - \frac{x'^2 + y'^2}{\lambda_p F^2} + \frac{u^2}{\lambda_f} \right] z. \tag{15}$$

To simplify the integral (11), we use the following cylindrical coordinates:

$$\begin{cases} x = \rho \cos \phi, & B_x(z) = B(z) \cos \alpha(z) \\ y = \rho \sin \phi, & B_y(z) = B(z) \sin \alpha(z) \end{cases} \quad (16)$$

with the inverse transformations given by

$$\begin{cases} \rho = \sqrt{x^2 + y^2}, & B(z) = \sqrt{B_x(z)^2 + B_y(z)^2} \\ \tan \phi = y/x, & \tan \alpha(z) = B_y(z)/B_x(z) \end{cases} \quad (17)$$

where the sign of the inverse tangent is taken to conform with the quadrant of x, y, and $B_x(z)$, $B_y(z)$, respectively. Equation (11) then becomes

$$\tilde{E}_d(\mathbf{r}') = \int_{-L/2}^{L/2} \exp\{i\pi C(z)\} \int_0^R \exp\{i\pi A(z)\rho^2\}$$
$$\cdot \int_{-\pi}^{\pi} \exp\{-i2\pi B(z)\rho \cos(\phi - \alpha(z))\} d\phi \rho d\rho dz.$$
$$(18)$$

The result for the innermost integral is well known, expressed in terms of the zero-order Bessel function of the first kind as

$$\int_{-\pi}^{\pi} \exp\{-i2\pi B(z)\rho \cos(\phi - \alpha(z))\} \, d\phi$$
$$= 2\pi J_0(2\pi B(z)\rho). \quad (19)$$

The next-level integral occurs in the calculation of the 3-D PSF of a lens near focus, and is written as

$$\int_0^1 \exp\left\{-\frac{i}{2}u\rho^2\right\} J_0(v\rho)\rho \, d\rho = \mathcal{L}(u, v) \quad (20)$$

where the real and imaginary parts of the function $\mathcal{L}(u,v)$ are expressed in terms of the Lommel functions. For more details, the reader may consult [130, Section 8.8, pp. 435–449]. In terms of the $\mathcal{L}$ function, the diffracted field at the detector is expressed as

$$\tilde{E}_d(\mathbf{r}') = 2\pi R^2 \int_{-L/2}^{L/2} \exp\{i\pi C(z)\}$$
$$\cdot \mathcal{L}(2\pi A(z)R^2, 2\pi B(z)R) \, dz. \quad (21)$$

The last integral is calculated numerically. Some properties of the volume hologram are now apparent qualitatively.

1) If the hologram is reconstructed at the recording wavelength ($\lambda_p = \lambda_f$), with a probe source at the same location as the reference source ($\mathbf{r}_p = \mathbf{r}_f$), and the detector is placed at the maximum of the Fourier transform of the signal ($x'/F = u$, $y'/F = 0$), then all the exponents in (18) vanish, and the reconstructed power is maximum. This condition is known as Bragg matching.

2) If either the reconstruction wavelength or the probe location change, the detector point $x'/F = u$, $y'/F = 0$ does not receive maximum power anymore. If the power drops uniformly over the entire detector plane, we say that the hologram is Bragg mismatched. When λ_p and $\mathbf{r}_p$ satisfy certain conditions, though, then
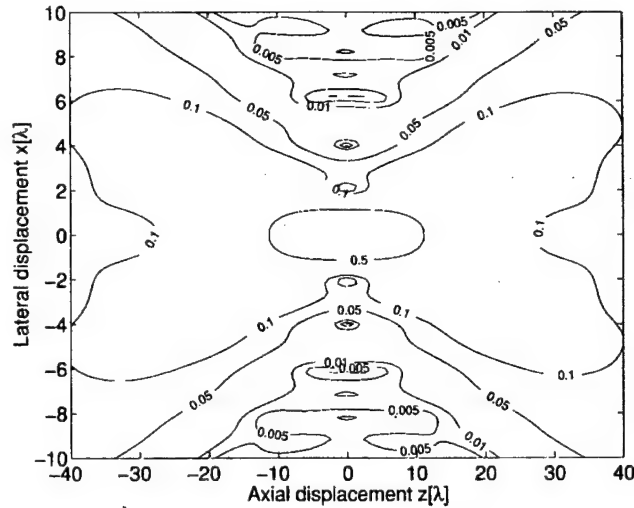
Fig. 11. Contour plot of the diffracted intensity measured by a detector at the focal point $x'/F = y'/F = 0$ of the geometry of Fig. 10, from a point source at position $\mathbf{r}_p = (x_p, 0, z_p)$ illuminating a volume hologram with $R = 750\lambda$, $L = 1500\lambda$. The same diffraction contour plot is obtained also from the geometries of Figs. 13 and 14.

significant power is diffracted into some other point on the detector plane ($x'/F \neq u$ or $y'/F \neq 0$). The locus of (λ_p, $\mathbf{r}_p$) over which this is possible is the degeneracy surface of the volume hologram. This effect and how it can be used to extract tomographic slices of polychromatic volume objects are the topics of Section IV-C.

3) Since $\mathcal{L}(\cdot, \cdot)$ describes the amplitude transmitted from a quadratic lens also, our result (21) shows that the diffracted light from the volume hologram is the coherent superposition of several "lenses" stacked in the $\hat{\mathbf{z}}$ direction. If the probe source is at the common front focus of all these virtual "lenses," then the "lenses" are all in phase and give a strong reconstruction in the back focal point (Bragg matched case). If the probe moves around or changes its color, the "lens" contributions will in general be out of phase (Bragg-mismatched case), except if the combination of probe position and wavelength and observation position are arranged such that the "lens stack" contributions are again in phase (degeneracy case).

The diffracted power received by a fixed detector pixel ($x'/F, y'/F$) when $\lambda_p = \lambda_f$, $y_p = 0$ are kept fixed and x_p, z_p are allowed to vary are plotted in Fig. 11. This response, which is common to other recording geometries as well (see below), should be compared with Fig. 12, which is the transmitted intensity captured by the detector if the volume hologram is replaced by a lens of the same aperture as function of x_p, z_p. The comparison explains, e.g., why the volume hologram is more efficient as a collector in a confocal microscope arrangement (Section III-C).

Now consider the case of a reflection hologram recorded with a spherical wave signal beam. The geometry is drawn in Fig. 13. The signal beam is now a spherical wave coun-
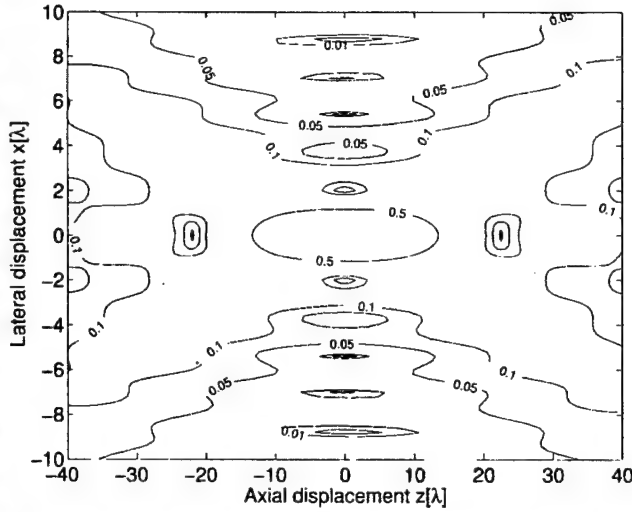
Fig. 12. Contour plot of the transmitted intensity, measured by a detector at the focal point of a thin quadratic lens with aperture $R = 750\lambda$, when illuminated from a point source at position $\mathbf{r}_p = (x_p, 0, z_p)$. This plot is, within scaling factors, the same as [130, Fig. 8.41] and is provided for comparison with Fig. 11.

terpropagating with respect to the reference, and coming to a focus at $\mathbf{r}_s = x_s\hat{\mathbf{x}} + y_s\hat{\mathbf{y}} + z_s\hat{\mathbf{z}}$. The expression of the electric field for this wave is

$$E_s(\mathbf{r}) = \exp\left\{i2\pi\frac{z - z_s}{\lambda_f} + i\pi\frac{(x - x_s)^2 + (y - y_s)^2}{\lambda_f(z - z_s)}\right\}. \tag{22}$$

To find the diffracted field, we start with an expression similar to (9) with (22) substituted in the refractive index modulation $\Delta\epsilon(\mathbf{r}) = E_f^*(\mathbf{r})E_s(\mathbf{r})$ and proceed as in the plane-wave signal case, omitting the Fourier-transform step since there is no lens in the arrangement of Fig. 13. The result is

$$E_d(\mathbf{r}') = 2\pi R^2 \int_{-L/2}^{L/2} \exp\left\{i\pi C(z)\right\}$$
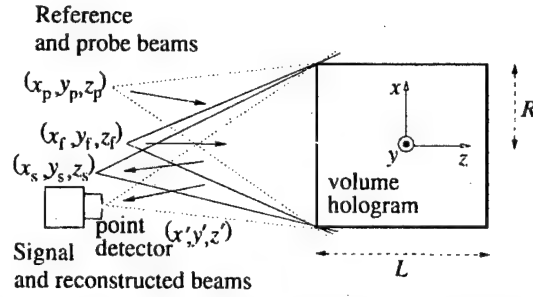$$\cdot \mathcal{L}(2\pi A(z)R^2, 2\pi B(z)R)\,dz \tag{23}$$

i.e., identical to (21), but with

$$A(z) = \frac{1}{\lambda_p(z - z_p)} + \frac{1}{\lambda_p(z - z')} - \frac{1}{\lambda_f(z - z_f)}$$
$$- \frac{1}{\lambda_f(z - z_s)} \tag{24}$$

$$B_x(z) = -\frac{x_p}{\lambda_p(z - z_p)} - \frac{x'}{\lambda_p(z - z')} + \frac{x_f}{\lambda_f(z - z_f)}$$
$$+ \frac{x_s}{\lambda_f(z - z_s)} \tag{25}$$

$$B_y(z) = -\frac{y_p}{\lambda_p(z - z_p)} - \frac{y'}{\lambda_p(z - z')} + \frac{y_f}{\lambda_f(z - z_f)}$$
$$+ \frac{y_s}{\lambda_f(z - z_s)} \tag{26}$$

$$C(z) = \frac{x_p^2 + y_p^2}{\lambda_f(z - z_p)} + \frac{x'^2 + y'^2}{\lambda_p(z - z')} - \frac{x_f^2 + y_f^2}{\lambda_f(z - z_f)}$$
$$- \frac{x_s^2 + y_s^2}{\lambda_f(z - z_s)} + 4\left(\frac{1}{\lambda_p} - \frac{1}{\lambda_f}\right)z. \tag{27}$$

Fig. 13. Schematic of the reflection geometry with spherical wave reference and signal beams.

The plot of diffracted power when λ_p, $\mathbf{r}'$ and y_p are fixed, while x_p, z_p vary, is virtually identical to that of Fig. 11 (this can be verified by comparing the expressions for the coefficients and the changes in the exponents as x_p, z_p change), so it will not be given again. The degeneracy surface for this geometry will be calculated in Section IV-C along with the other recording geometries.

B. 90° Geometry Holograms

The 90° geometry differs from the reflection geometry because the paraxial approximation for the signal and diffracted beams is made along the $\hat{\mathbf{x}}$ rather than the $\hat{\mathbf{z}}$ axis. This leads to quantitatively different expressions in the diffraction integrals. We will examine in this section the case of a plane wave signal only, incident at angle $u \ll 1$ with respect to the $\hat{\mathbf{x}}$ axis, as shown in Fig. 14. The reference and signal beams are

$$E_f(\mathbf{r}) = \exp\left\{i2\pi\frac{z - z_f}{\lambda_f} + i\pi\frac{(x - x_f)^2 + (y - y_f)^2}{\lambda_f(z - z_f)}\right\} \tag{28}$$

$$E_s(\mathbf{r}) = \exp\left\{-i2\pi\left(1 - \frac{u^2}{2}\right)\frac{x}{\lambda} + i2\pi u\frac{z}{\lambda}\right\}. \tag{29}$$

For a probe field

$$E_p(\mathbf{r}) = \exp\left\{i2\pi\frac{z - z_p}{\lambda_p} + i\pi\frac{(x - x_p)^2 + (y - y_p)^2}{\lambda_p(z - z_p)}\right\} \tag{30}$$

and an observation point $\mathbf{r}''$ near the $\hat{\mathbf{x}}$ axis, the diffraction integral in the paraxial approximation is given as

$$E_d(\mathbf{r}'') = \iiint E_p(\mathbf{r})\Delta\epsilon(\mathbf{r})\,\text{circ}\left(\frac{\sqrt{x^2 + y^2}}{R}\right)\text{rect}\left(\frac{z}{L}\right)$$
$$\times \exp\left\{i2\pi\frac{x'' - x}{\lambda_p}\right.$$
$$\left. + i\pi\frac{(y'' - y)^2 + (z'' - z)^2}{\lambda_p(x'' - x)}\right\}d^3\mathbf{r}. \tag{31}$$

Performing the Fourier transform in y'', z'', and subsequently the integrals in cylindrical coordinates as in Section IV-A for the plane wave case, we obtain yet again a result of the form

$$\tilde{E}_d(\mathbf{r}') = 2\pi R^2 \int_{-L/2}^{L/2} \exp\left\{i\pi C(z)\right\}$$
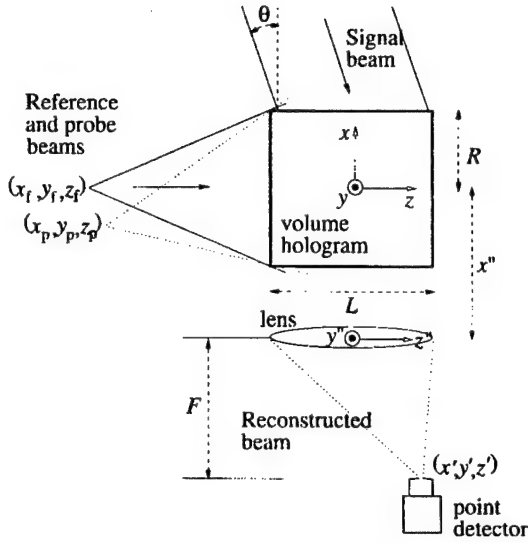$$\cdot \mathcal{L}(2\pi A(z)R^2, 2\pi B(z)R)\,dz. \tag{32}$$

Fig. 14. Schematic of the 90° geometry with spherical wave reference and plane wave signal beams.

This time the coefficients are given by

$$A(z) = \frac{1}{\lambda_p(z - z_p)} - \frac{1}{\lambda_f(z - z_f)} \qquad (33)$$

$$B_x(z) = -\frac{x_p}{\lambda_p(z - z_p)} + \frac{x_f}{\lambda_f(z - z_f)}$$
$$+ \frac{y'^2 + z'^2}{2\lambda_p F^2} - \frac{u^2}{2\lambda_f} + \frac{1}{\lambda_f} - \frac{1}{\lambda_p} \qquad (34)$$

$$B_y(z) = -\frac{y_p}{\lambda_p(z - z_p)} + \frac{y_f}{\lambda_f(z - z_f)} - \frac{y'}{\lambda_p F} \qquad (35)$$

$$C(z) = \frac{x_p^2 + y_p^2}{\lambda_f(z - z_p)} - \frac{x_f^2 + y_f^2}{\lambda_f(z - z_f)}$$
$$+ 2\left(\frac{u}{\lambda_f} - \frac{z'}{\lambda_p F} + \frac{1}{\lambda_p} - \frac{1}{\lambda_f}\right)z. \qquad (36)$$

The change in diffracted power as function of x_p, z_p is, once again, identical to Fig. 11, and will not be given separately. The degeneracies will be calculated immediately below.

C. Hologram Degeneracies and Multispectral Tomography

We now turn to the calculation of the hologram degeneracies, i.e., the conditions for achieving significant reconstructed power even when the probe field is not a replica of the reference field. From the diffraction integrals (21), (23), (32), we can see that the condition for obtaining significant reconstruction is equivalent to setting the arguments of $\mathcal{L}(\cdot, \cdot)$ as well as the varying portion of $C(\cdot)$ equal to zero. If these conditions are not satisfied, the value of the integral decreases, i.e., Bragg mismatch occurs.

Obviously, in each geometry there are several parameters that one may manipulate in order to eliminate the exponents. The selection depends on the application. We are interested in the case of a reconstructing field produced by an extended polychromatic object, and a planar two-dimensional detector located at the exit plane of the system. Then, the parametrization of interest is the locus and

wavelength of the point radiators within the object that produce maximum reconstructed power on a particular pixel of the detector as a function of the pixel coordinates on the detector. We will see that this parametrization results, in each case, in a surface in object space; this is the degeneracy surface for our chosen planar detector geometry (nonplanar detector surfaces would yield different degeneracy surfaces but are hard to come about in practice). The reconstructing wavelength must vary across the degeneracy surface, too, for maximum reconstructed power. Thus, the field diffracted off the volume hologram "isolates" a surface subset of radiators in space, as well as filters them in color. This is equivalent to a tomographic slicing operation in both space and spectral domains. By scanning the volume hologram in two dimensions, the full four-dimensional (4-D) reconstruction of the object (in space and color) can be obtained. We now derive the degeneracy surface shapes for various holographic recording geometries, in order to demonstrate the operation of the volume hologram as a spatiospectral filter.

We begin with the case of a plane wave signal, reflection geometry hologram, as in the first part of Section IV-A. For later convenience, we define the parameter $\mu = \lambda_p/\lambda_f$. To derive the degeneracy surface, we set all coefficients $A(z)$, $B_x(z)$, $B_y(z)$, $C(z)$ equal to zero, at least to first order in z. From (12) we obtain

$$z_p = \frac{z_f}{\mu}. \qquad (37)$$

From (13), and using (37), follows

$$x_p = x_f + z_f\left(\frac{1}{\mu}\frac{x'}{F} - u\right) \qquad (38)$$

whereas (14) and (37) yield

$$y_p = y_f + \frac{z_f}{\mu}\frac{y'}{F}. \qquad (39)$$

Substituting into (15) results in the following quadratic equation in μ

$$D\mu^2 + G\mu - H = 0 \qquad (40)$$

where

$$D = \left(\frac{x_f}{z_f} - u\right)^2 + \frac{y_f^2}{z_f^2} \qquad (41)$$

$$G = 4 - 2u\frac{x_f}{z_f} + \frac{x'^2 + y'^2}{F^2} - \left(\frac{x_f}{z_f} - u - \frac{x'}{F}\right)^2$$
$$- \left(\frac{y_f}{z_f} - \frac{y'}{F}\right)^2 \qquad (42)$$

$$H = 4 - 2\frac{x'^2 + y'^2}{F^2}. \qquad (43)$$

Therefore, the degeneracy surface is obtained by setting μ equal to the root of (40) that is closest to 1 in magnitude and then substituting in (37)–(39). The result for a particular
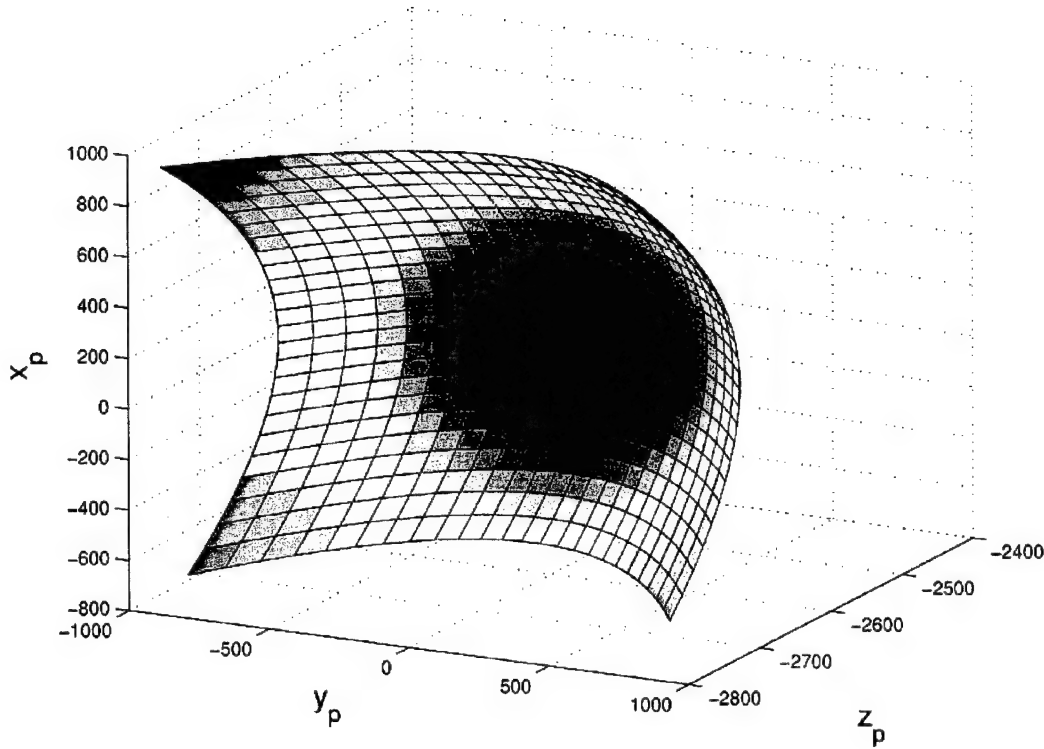
Fig. 15. Degeneracy surface (space and color) of the reflection recording geometry with a plane wave signal, computed numerically. The parameters for this plot were $\mathbf{r}_f = (100, -100, -2500)\lambda$, $u = 0, -0.3 \leq x'/F, y'/F \leq 0.3$. In the plot, the blue color corresponds to $\mu = 0.899$ and the purple to $\mu = 1.0$.

numerical example is shown in Fig. 15. Two important points about the surface of Fig. 15 should be noted.

1) The degeneracy surface is not infinitely thin as implied by the drawing but has a finite thickness because the reconstructed intensity from points and colors near the surface is not zero but falls off smoothly according to (21).
2) The diffraction efficiency from points belonging to the degeneracy surface is not uniformly 1 but falls off toward the surface edges because higher order z terms in the exponents cause weak Bragg mismatch; this deviation from true degeneracy is also calculated by use of (21).

Numerical results for the surface thickness and deviation from degeneracy are not given here but can be calculated easily. These remarks also hold for the surfaces computed later in this section.

The derivation of the degeneracy surfaces for other recording geometries is similar to the one just described and will not be given here; only the results will be quoted.

For the spherical wave signal reflection geometry case, μ is the root, closest in magnitude to 1, of

$$D\mu^2 + G\mu - H = 0 \tag{44}$$

where

$$D = \left(\frac{x_f}{z_f} + \frac{x_s}{z_s}\right)^2 + \left(\frac{y_f}{z_f} + \frac{y_s}{z_s}\right)^2 \tag{45}$$

$$G = 4 - 2\frac{x_f x_s + y_f y_s}{z_f} + \frac{x'^2 + y'^2}{z'^2}$$
$$- \left(\frac{x_f}{z_f} + \frac{x_s}{z_s} + \frac{x'}{F}\right)^2 - \left(\frac{y_f}{z_f} + \frac{y_s}{z_s} + \frac{y'}{F}\right)^2 \tag{46}$$

$$H = 4 - 2\frac{x'^2 + y'^2}{F^2}. \tag{47}$$

The spatial coordinates of the degeneracy surface are obtained from

$$\frac{1}{z_p} = \mu\left(\frac{1}{z_f} + \frac{1}{z_s}\right) - \frac{1}{z'} \tag{48}$$

$$\frac{x_p}{z_p} = \mu\left(\frac{x_f}{z_f} + \frac{x_s}{z_s}\right) - \frac{x'}{z'} \tag{49}$$

$$\frac{y_p}{z_p} = \mu\left(\frac{y_f}{z_f} + \frac{y_s}{z_s}\right) - \frac{y'}{z'}. \tag{50}$$

A numerical example is given in Fig. 16. It is interesting to note that the degeneracy surface in this case is identical to the surface obtained in the case of a plane wave signal (derived immediately above) with $x_s = -u z_s$, $y_s = 0$, and inverted output coordinates ($x' \leftrightarrow -x'$, $y' \leftrightarrow -y'$).

For the plane wave signal 90° geometry case, the quadratic equation for μ is also of the form

$$D\mu^2 + G\mu - H = 0 \tag{51}$$

where

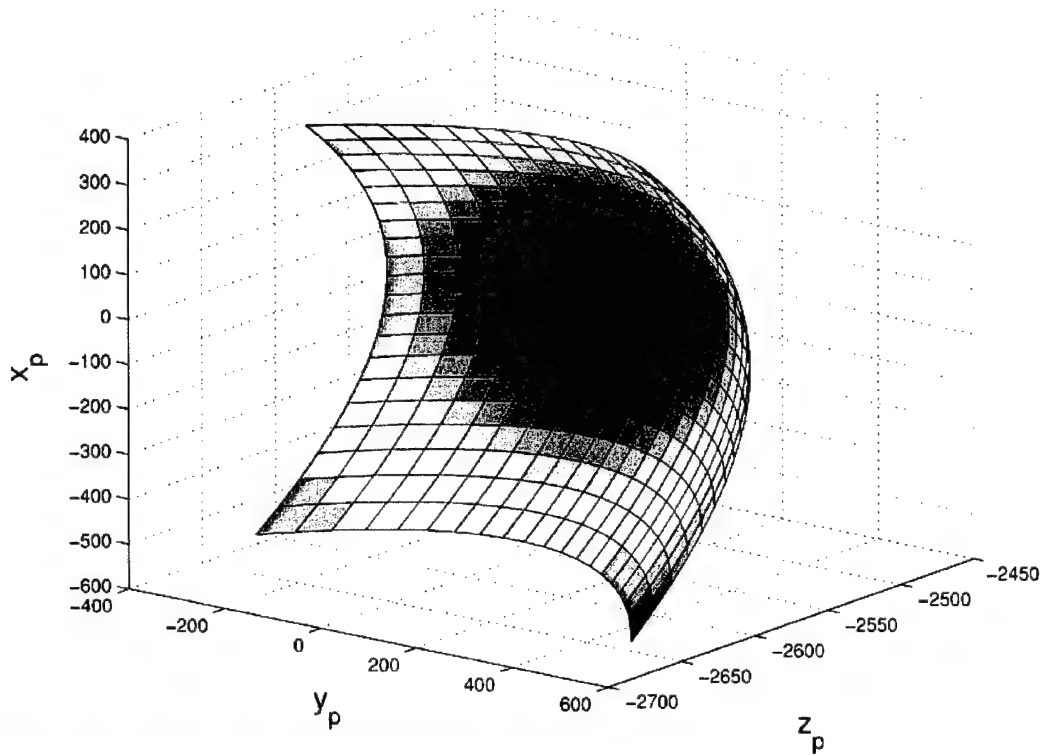$$D = 1 + \frac{x_f^2 + y_f^2}{z_f^2} - 2\frac{x_f}{z_f} - u^2 \tag{52}$$

Fig. 16. Degeneracy surface (space and color) of the reflection recording geometry with a spherical wave signal, computed numerically. The parameters for this plot were $\mathbf{r}_f = (-100, 100, -2500)\lambda$, $\mathbf{r}_s = (0, 0, -3000)\lambda$, $-500\lambda \leq x', y' \leq 500\lambda, z' = -3000\lambda$. In the plot, the blue color corresponds to $\mu = 0.966$ and the purple to $\mu = 1.0$.

$$G = \left(1 + \frac{x_f}{z_f}\right)^2 + \left(\frac{y_f}{z_f} + \frac{y'}{F}\right)^2 + \frac{z'^2}{F^2}$$
$$+ (u-1)^2 - 2\left(1 + \frac{x_f^2 + y_f^2}{z_f^2}\right) \quad (53)$$

$$H = \left(1 - \frac{z'}{F}\right)^2. \quad (54)$$

After solving for μ as the root of the quadratic with magnitude closest to 1, the spatial coordinates are obtained from

$$z_p = \frac{z_f}{\mu} \quad (55)$$

$$\frac{x_p}{z_p} = -\mu\left(1 - \frac{x_f}{z_f} - \frac{u^2}{2}\right) + \left(1 - \frac{y'^2 + z'^2}{2F^2}\right) \quad (56)$$

$$\frac{y_p}{z_p} = \mu\frac{y_f}{z_f} + \frac{y'}{F}. \quad (57)$$

A numerical example is given in Fig. 17.

For the sake of completeness, we also give the result for the degeneracy surface of a volume hologram recorded with a spherical wave signal in the 90° geometry (see Fig. 18 for the notation). The calculation of the diffracted field cannot be done under the framework of Sections IV-A and IV-B and will not be given here. However, the degeneracy derivation is straightforward. It results also in a quadratic equation for μ of the form

$$D\mu^2 + G\mu - H = 0 \quad (58)$$

where

$$D = 1 + \frac{x_f^2 + y_f^2}{z_f^2} - 2\left(\frac{x_f}{z_f} + \frac{y_f}{z_f}\frac{y_s}{x_s}\right) - \frac{z_s^2}{x_s^2} \quad (59)$$

$$G = \left(1 + \frac{x_f}{z_f}\right)^2 + \left(\frac{y_f}{z_f} + \frac{y'}{x'}\right)^2 + \frac{z'^2}{x'^2}$$
$$+ \frac{y_s^2 + z_s^2}{x_s^2} - 1 - 2\frac{x_f^2 + y_f^2}{z_f^2} - 2\frac{y'y_s}{x'x_s} \quad (60)$$

$$H = 1 + \frac{z'^2}{x'^2}. \quad (61)$$

The spatial coordinates of the degeneracy surface are given by

$$z_p = \frac{z_f}{\mu} \quad (62)$$

$$\frac{x_p}{z_p} = -\mu\left(1 - \frac{x_f}{z_f} - \frac{y_s^2 + z_s^2}{2x_s^2}\right) + \left(1 - \frac{y'^2 + z'^2}{2x'^2}\right) \quad (63)$$

$$\frac{y_p}{z_p} = \mu\left(\frac{y_f}{z_f} - \frac{y_s}{x_s}\right) + \frac{y'}{x'} \quad (64)$$

with μ computed from (58). A numerical example is given in Fig. 19.

V. STATISTICAL PROPERTIES OF DIFFRACTION FROM VOLUME GRATINGS

In the examples of the previous sections, we made assumptions about the nature of the random objects reconstructing the volume hologram, as well as about the form of the hologram itself. In particular, we worked with volume
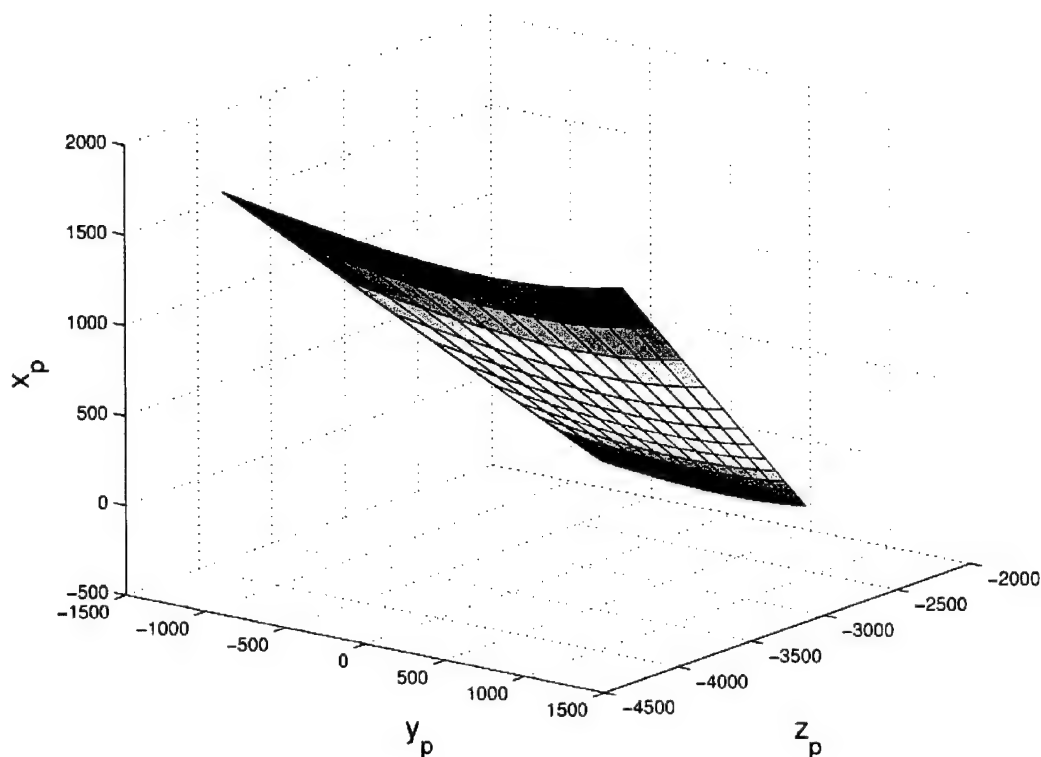
Fig. 17. Degeneracy surface (space and color) of the 90° recording geometry with a plane wave signal, computed numerically. The parameters for this plot were $\mathbf{r}_f = (100, -100, -2500)\lambda$, $u = 0, -0.3 \leq x'/F, y'/F \leq 0.3$. In the plot, the blue color corresponds to $\mu = 0.619$ and the purple to $\mu = 1.246$.

holograms recorded with a spherical reference beam and derived their operation as point-source correlators for imaging under two conditions: 1) when replacing the collector lens in a fluorescent confocal microscope with a monochromatic object (Section III-C), they improve the resolution of the regular confocal arrangement, and 2) when reconstructing a polychromatic (4-D) object, they isolate a surface in the space and wavelength domain, allowing the full 4-D tomographic reconstruction of the object with appropriate scanning (Section IV-C). Color-selective tomography is a unique property of volume holograms as optical elements and cannot be achieved by a design that incorporates planar optical elements only. In this section, we generalize the volume hologram operation as shaping the modes of the object field through correlation (mixing) with the modes of the volume hologram.

As we mentioned already in Section III-A, computational imaging is performed, in general, by transformations on the intensity values of the field at the detector plane. When viewed as an analysis problem, this means that one needs to know the statistical properties of the intensity values at the output plane as a function of the imaging system and the statistics of the object. From the design point of view, it is desired to construct the imaging system so as to shape the output intensity statistics appropriately for the task at hand. The intensity transformation law between the source intensity distribution $I_s(\mathbf{r}')$ and the intensity distribution at the detector space $I_d(\mathbf{r})$ is, in the case of a completely
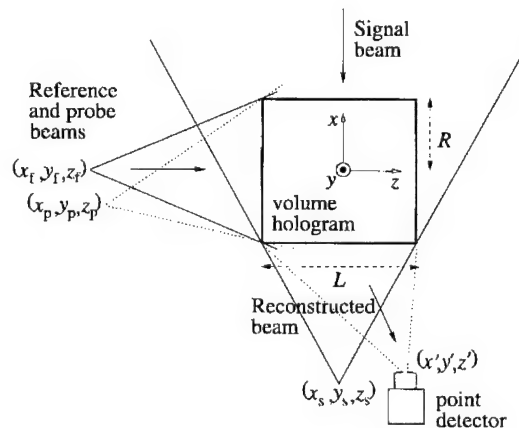
Fig. 18. Schematic of the 90° geometry with sperical wave reference and signal beams.

incoherent source, of the form

$$I_d(\mathbf{r}') = \int_{V_S} I_s(\mathbf{r}) h(\mathbf{r}', \mathbf{r}) \, d^3\mathbf{r} \qquad (65)$$

where the transfer function $h(\mathbf{r}', \mathbf{r})$ describes the operation of the optical system.

In many imaging systems, there is a single optical beam propagating between the object and image. Such was the operation of the volume holographic system of Figs. 6, 10, 13, 14, and 18. We would like to characterize the operation of a volume hologram in configurations where two or more beams interfere on the detector plane. In such
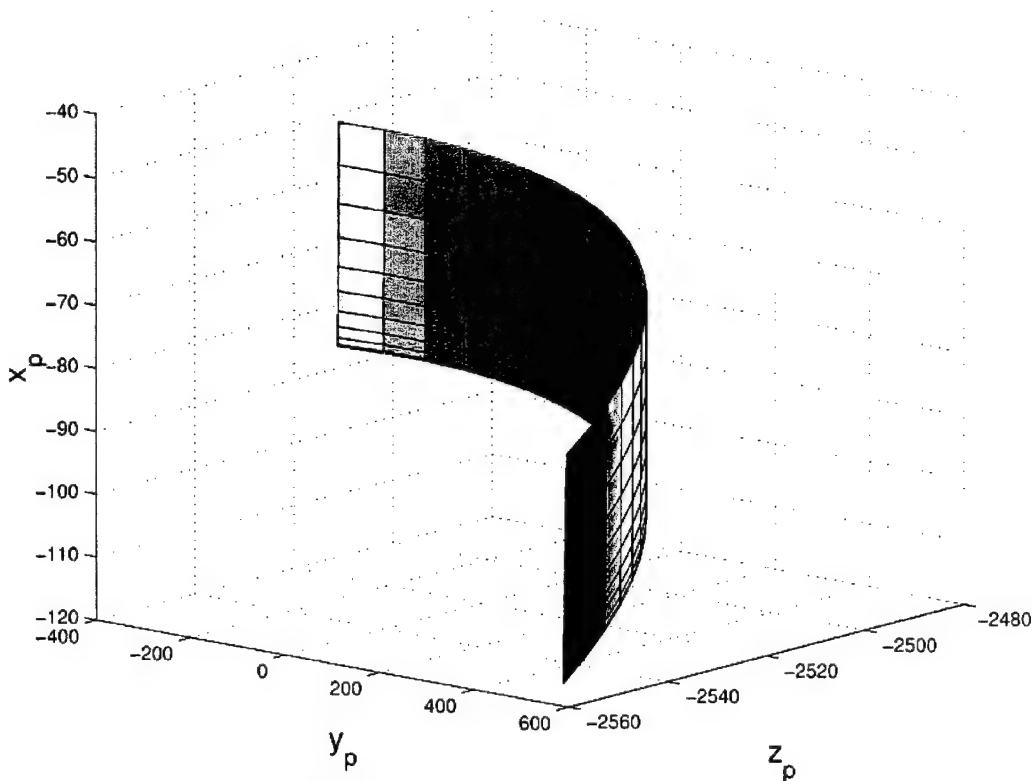
Fig. 19. Degeneracy surface (space and color) of the 90° recording geometry with a spherical wave signal, computed numerically. The parameters for this plot were $\mathbf{r}_f = (-100, 100, -2500)\lambda$, $\mathbf{r}_s = (3000, 0, 0)\lambda$, $-500\lambda \le y', z' \le 500\lambda, x' = 3000\lambda$. In the plot, the blue color corresponds to $\mu = 0.979$ and the purple to $\mu = 1.0$.

interferometric systems, the $I_d(\mathbf{r}')$ measurement contains information about the statistical correlation properties of the field at the output. We therefore need to model the effect of volume diffraction on the correlation properties of a random optical field. To this end, in the remainder of this section, we first introduce the notation and terminology of statistical optics [131], [132] and rederive the deterministic correlation property of volume diffraction; we then generalize the correlation property in a statistical framework and use the modal decomposition of both the field coherence function and the hologram refractive index modulation in order to formulate the operation of the volume hologram as a design problem.

Suppose the geometry of an imaging system is such that the fields from two observation points $\mathbf{r}'_1(\mathbf{r}')$, $\mathbf{r}'_2(\mathbf{r}')$ interfere at every point $\mathbf{r}'$ in the detector space. The measurement then is (dropping the $\mathbf{r}'$ dependence of $\mathbf{r}'_1$, $\mathbf{r}'_2$ for notational simplicity)

$$I_d(\mathbf{r}') = \text{E.V.}\{|E_d(\mathbf{r}'_1) + E_d(\mathbf{r}'_2)|^2\}. \quad (66)$$

Using the definitions

$$I_d(\mathbf{r}'_j) = \text{E.V.}\{|E_d(\mathbf{r}'_j)|^2\} \qquad j = 1, 2 \quad (67)$$

$$\Gamma_d(\mathbf{r}'_1, \mathbf{r}'_2, \tau) = \text{E.V.}\{E^*(\mathbf{r}'_1, t)E(\mathbf{r}'_2, t + \tau)\} \quad (68)$$

where τ is the relative time delay between the two optical paths and Γ_d is the second-order correlation function of the random optical field, we obtain for the detected intensity

the result

$$I_d(\mathbf{r}') = I_d(\mathbf{r}'_1) + I_d(\mathbf{r}'_2) + 2 \,\text{Re}\, \Gamma_d(\mathbf{r}'_1, \mathbf{r}'_2, \tau). \quad (69)$$

The last result states that the interferometric measurement contains the field correlation information superimposed on a quasi-uniform bright background (typically, the variation of $I_d(\mathbf{r}'_1)$ with $\mathbf{r}'_1$ is very small). Instead of $\Gamma_d(\mathbf{r}'_1, \mathbf{r}'_2, \tau)$, it is typical to assume that the field is quasi-monochromatic, and that the path delay is $\tau = 0$, and use the mutual intensity $J_d(\mathbf{r}'_1, \mathbf{r}'_2)$, defined as

$$J_d(\mathbf{r}'_1, \mathbf{r}'_2) = \Gamma_d(\mathbf{r}'_1, \mathbf{r}'_2, 0). \quad (70)$$

Relation (69) then becomes

$$I_d(\mathbf{r}') = I_d(\mathbf{r}'_1) + I_d(\mathbf{r}'_2) + 2 \,\text{Re}\, J_d(\mathbf{r}'_1, \mathbf{r}'_2). \quad (71)$$

These results are well known and hold for any interferometric optical system; we seek to compute the effect of a volume hologram on the mutual intensity $J_d(\mathbf{r}'_1, \mathbf{r}'_2)$. Before getting to that point, though, it is useful to repeat a deterministic property of volume diffraction: if a volume hologram is illuminated by a complex (coherent) field, then the diffracted field contains the 2-D correlation between the input field and the pattern(s) stored in the modulated refractive index of the hologram. The generalization to statistical fields is then obvious.

Let V_S be the volume where a three-dimensional monochromatic (at wavelength $\lambda = 2\pi/k$) probe light source is confined, and $E_p(\mathbf{r}'')$ the field emitted by an

infinitesimal source volume at $\mathbf{r}'' \in V_S$. Let $V_{\mathcal{H}}$ be the volume where a 3-D perturbation $\Delta\epsilon(\mathbf{r})$ of the refractive index is confined ($\Delta\epsilon(\mathbf{r}) = 0$ for $\mathbf{r} \notin V_{\mathcal{H}}$), i.e., a volume hologram. When the field scattered by V_S reaches $V_{\mathcal{H}}$, a secondary diffracted field $E_d(\mathbf{r}')$ ($\mathbf{r}' \notin V_S, V_{\mathcal{H}}$) is produced. Let $\hat{\mathbf{n}}$ denote a vector of unit magnitude in an arbitrary direction that from here on we refer to as "optical axis." The 3-D spatial spectrum of the source field is denoted as

$$\tilde{E}_s(\mathbf{k}) = \int_{\mathbf{R}^3} E_s(\mathbf{r}) \exp\{-i\mathbf{k} \cdot \mathbf{r}\} \, d^3\mathbf{r} \qquad (72)$$

with a similar expression for $\tilde{E}_d$. We also denote

$$\mathbf{k}_\perp = \mathbf{k} \times \hat{\mathbf{n}}, \quad k_\| = \mathbf{k} \cdot \hat{\mathbf{n}} \qquad \left(|\mathbf{k}_\perp|^2 + k_\|^2 = k^2\right).$$

The correlation property is then expressed as follows:

$$\tilde{E}_d(\mathbf{k}) = \frac{1}{k_\|} \left(\frac{\tilde{E}_s}{k_\|} \star \Delta\tilde{\epsilon}^* \right) \qquad (73)$$

for $\mathbf{k}$ satisfying $|\mathbf{k}| = k$, where the $\star$ denotes correlation and $\Delta\tilde{\epsilon}^*$ is the complex conjugate of $\Delta\tilde{\epsilon}$. The proof of this statement is given in Appendix I.

The generalization of (73) to a random field is straightforward. We define the spatial Fourier transform of the source mutual intensity function as

$$\tilde{J}_s(\mathbf{k}_1, \mathbf{k}_2) = \iint_{\mathbf{R}^2} J_s(\mathbf{r}_1, \mathbf{r}_2)$$
$$\cdot \exp\{-i(\mathbf{k}_1 \cdot \mathbf{r}_1 + \mathbf{k}_2 \cdot \mathbf{r}_2)\} \, d^2\mathbf{r}_1 \, d^2\mathbf{r}_2 \quad (74)$$

with a similar expression for $\tilde{J}_d(\mathbf{k}_1, \mathbf{k}_2)$. Then, under the constraint $|\mathbf{k}_1| = |\mathbf{k}_2| = k$, $\tilde{J}_d(\mathbf{k}_1, \mathbf{k}_2)$ is related to $\tilde{J}_s(\mathbf{k}_1, \mathbf{k}_2)$ through

$$\tilde{J}_d(\mathbf{k}_1, \mathbf{k}_2) = \iint_{\mathbf{R}^2} \frac{\tilde{J}_s(\boldsymbol{\kappa}_1'', \boldsymbol{\kappa}_2'')}{k_{1,\|} k_{2,\|} k_{1,\|}'' k_{2,\|}''} \Delta\tilde{\epsilon}(\boldsymbol{\kappa}_1'' - \mathbf{k}_1)$$
$$\cdot \Delta\tilde{\epsilon}^*(\boldsymbol{\kappa}_2'' - \mathbf{k}_2) \, d^2\mathbf{k}_{1,\perp} \, d^2\mathbf{k}_{2,\perp}. \quad (75)$$

Similar to Appendix I, we used the notation $\boldsymbol{\kappa}_j'' = \mathbf{k}_{j,\perp} + \hat{\mathbf{n}}\sqrt{k^2 - |\mathbf{k}_{j,\perp}|^2}$, $j = 1, 2$. The proof of (75) is given in Appendix II. (This expression also follows directly by specializing a result derived in [112] to the case of a volume hologram; the full proof is given here for completeness.) This is also a correlation relation between the six-dimensional functions $\tilde{J}_d(\mathbf{k}_1, \mathbf{k}_2)$ and $\Delta\tilde{\epsilon}^*(\mathbf{k}_1)\Delta\tilde{\epsilon}(\mathbf{k}_2)$ constrained on the 4-D sphere $|\mathbf{k}_1| = |\mathbf{k}_2| = k$.

We now seek to cast (75) as a design problem. To this end, we decompose both the Fourier-transformed mutual intensities $\tilde{J}_d(\mathbf{k}_1, \mathbf{k}_2)$, $\tilde{J}_s(\mathbf{k}_1, \mathbf{k}_2)$ and the index modulation $\Delta\tilde{\epsilon}(\mathbf{k})$ in their respective modes and show how the modes mix as a result of volume diffraction.

The coherent mode decomposition property for a general random field states that, under some general existence and continuity conditions for the cross-spectral density, the quantity $J(\mathbf{r}_1, \mathbf{r}_2)$ can be decomposed into a sum of products of orthogonal modes. For the source field, the decomposition is written as follows:

$$J_s(\mathbf{r}_1, \mathbf{r}_2) = \sum_\mathbf{m} \alpha_\mathbf{m} \psi_\mathbf{m}^*(\mathbf{r}_1)\psi_\mathbf{m}(\mathbf{r}_2) \qquad (76)$$

where the $\psi_\mathbf{m}$ are the eigenfunctions and the $\alpha_\mathbf{m}$ the eigenvalues of the Friedholm-type integral equation

$$\int_{V_S} J_s(\mathbf{r}_1, \mathbf{r}_2)\psi_\mathbf{m}(\mathbf{r}_2) \, d^3\mathbf{r}_1 = \alpha_\mathbf{m}\psi_\mathbf{m}(\mathbf{r}_1). \qquad (77)$$

The orthogonality of the eigenfunctions $\psi_\mathbf{m}$ is expressed by

$$\int_{V_S} \psi_\mathbf{m}^*(\mathbf{r})\psi_{\mathbf{m}'}(\mathbf{r}) \, d^3\mathbf{r} = \delta_{\mathbf{mm}'}. \qquad (78)$$

It is straightforward to prove that the decomposition also holds in the spatial-spectral domain; i.e.,

$$\tilde{J}_s(\mathbf{k}_1, \mathbf{k}_2) = \sum_\mathbf{m} \alpha_\mathbf{m} \tilde{\psi}_\mathbf{m}(\mathbf{k}_1)\tilde{\psi}_\mathbf{m}(\mathbf{k}_2). \qquad (79)$$

Since the volume $V_{\mathcal{H}}$ occupied by the hologram is finite, we can decompose $\Delta\epsilon(\mathbf{r})$ into a Fourier series, according to

$$\Delta\tilde{\epsilon}(\mathbf{k}) = \sum_\mathbf{j} \epsilon_\mathbf{j}\tilde{\phi}_\mathbf{j}(\mathbf{k}), \quad \epsilon_\mathbf{j} = \int_{V_{\mathcal{H}}} \Delta\epsilon(\mathbf{r})\phi_\mathbf{j}(\mathbf{r}) \, d^3\mathbf{r}. \quad (80)$$

The orthogonality condition for the basis $\{\phi_\mathbf{j}\}_{\mathbf{j} \in \mathbb{N}^3}$ is expressed as

$$\int_{V_{\mathcal{H}}} \phi_\mathbf{j}^*(\mathbf{r})\phi_{\mathbf{j}'}(\mathbf{r}) \, d^2\mathbf{r} = \delta_{\mathbf{jj}'}. \qquad (81)$$

We seek to compute the coherent mode decomposition of the mutual intensity of the diffracted field in terms of the modes (76) of the source mutual intensity and the modes (80) of the volume hologram. In Appendix III we show that

$$\tilde{J}_d(\mathbf{k}_1, \mathbf{k}_2) = \sum_\mathbf{m} \alpha_\mathbf{m} \tilde{\varphi}_\mathbf{m}^*(\mathbf{k}_1)\tilde{\varphi}_\mathbf{m}(\mathbf{k}_2) \qquad (82)$$

where, for $|\mathbf{k}| = k$

$$\tilde{\varphi}_\mathbf{m}(\mathbf{k}) = \frac{1}{k_\|} \sum_\mathbf{j} \epsilon_\mathbf{j} \int_{\mathbf{R}^2} \tilde{\psi}_\mathbf{m}^*(\mathbf{k}')\tilde{\phi}_\mathbf{j}(\mathbf{k}' - \mathbf{k})\frac{d^2\mathbf{k}_\perp'}{k_\|'}. \quad (83)$$

In the space domain, the above expression can be rewritten as

$$\varphi_\mathbf{m}(\mathbf{r}) = \sum_\mathbf{j} \epsilon_\mathbf{j} \int_{V_S} \int_{V_{\mathcal{H}}} \psi_\mathbf{m}^*(\mathbf{r}'')\phi_\mathbf{j}(\mathbf{r}')$$
$$\cdot \frac{\exp\{ik(|\mathbf{r}'' - \mathbf{r}'| - |\mathbf{r}' - \mathbf{r}|)\}}{|\mathbf{r}'' - \mathbf{r}'||\mathbf{r}' - \mathbf{r}|} \, d^3\mathbf{r}' \, d^3\mathbf{r}''. \quad (84)$$

Expressions (83) and (84) are the key results of this section: they express the modal structure of the diffracted field as a mixture of the coherence modes of the source and the modes of the volume hologram [note that the modes in (83) or (84) are not orthogonal but can be orthonormalized in straightforward fashion with a Gramm–Schmidt procedure]. The mixing occurs primarily through the coupling constants $\epsilon_\mathbf{j}$, i.e., the Fourier components of the refractive index modulation, while the modes $\tilde{\phi}_\mathbf{j}$ of the hologram itself act as weighting functions. Thus, the design problem is

defined in terms of (82) and (83) or (84): specify the Fourier components ϵ_j such that the desired mutual coherence function $\tilde{J}_d(\mathbf{k}_1, \mathbf{k}_2)$ is synthesized.

We will not attempt examples of modal design in this paper, but will point out two conclusions that follow from our results.

1) In the examples of Sections III-C and IV, the field is correlated with the particular spherical-reference mode of the volume hologram. This action is similar to a matched filter. The response, described by the degeneracy surfaces, is the set of object field modes that correlate with the hologram mode.

2) The response of the volume hologram can become much richer simply by following the modal synthesis approach delineated before; this is because the coupling coefficients ϵ_j provide three degrees of design freedom (j is a three-element vector) thanks to the 3-D nature of the volume hologram.

VI. CONCLUSIONS AND DISCUSSION

In this paper we introduced the concept of using volume holograms for multidimensional imaging and demonstrated numerically various imaging functions that a volume hologram can perform. The specific geometries of confocal microscopy with volume holographic collector and color-selective tomography are of immediate interest, and we are currently working on experimental demonstrations. The modal approach outlined in Section V extends optical engineering design from the traditional surface-to-surface transformations to the most general domain of volume transformations. This "3-D optical engineering" approach is well tuned to the construction of hybrid optical systems, where optics perform analog transformations at the front end, while back-plane digital electronic computations provide the transformations that optical elements cannot do well (e.g., Fourier transforms on the intensity function, nonlinearities, etc.) thus completing the generality of the system. Including volume holograms as analog optical elements in the design permits maximum flexibility in the quest for the optimal system.

The commercial value of ubiquitous imaging will undoubtedly increase rapidly with the ongoing revolutions of digital and hybrid imaging. Humans are known to be "visual" animals, i.e., in most situations they respond optimally to visual stimulation. In many instances, sophisticated visual interfaces can drastically improve the performance of critical social functions as diverse as education of young children and national or corporate security. Advanced interfaces are also necessary in the domains of machine vision and machine learning for the improvement of algorithms or even the invention of new ones based on the availability of more complete visual information about the surrounding world. The endowment of optical systems with powerful elements, such as volume holograms, and new design approaches geared toward advanced imaging and visual interface is critical for the achievement of these technological advances in the near future.

APPENDIX I
PROOF OF THE CORRELATION PROPERTY FOR THE DIFFRACTION OF A COMPLEX FIELD FROM A VOLUME HOLOGRAM

We will now prove assertion (73). Let us denote by

$$E_i(\mathbf{r}) = \int_{V_S} E_s(\mathbf{r}'') \frac{\exp\{ikR''\}}{R''} \, d^3\mathbf{r}'' \tag{85}$$

the field incident on location $\mathbf{r} \in V_{\mathcal{H}}$ of the volume hologram, where $R'' = |\mathbf{r} - \mathbf{r}''|$. Using Born's approximation, the field diffracted from the volume hologram is given by

$$E_d(\mathbf{r}') = \int_{V_{\mathcal{H}}} E_i(\mathbf{r})\Delta\epsilon(\mathbf{r}) \frac{\exp\{ikR\}}{R} \, d^3\mathbf{r} \tag{86}$$

$$= \int_{V_S} \int_{V_{\mathcal{H}}} E_s(\mathbf{r}'')\Delta\epsilon(\mathbf{r}) \frac{\exp\{ikR''\}}{R''} \frac{\exp\{ikR\}}{R} \cdot d^3\mathbf{r} \, d^3\mathbf{r}'' \tag{87}$$

where $R = |\mathbf{r}' - \mathbf{r}|$. We now use Weyl's identity for the expansion of a point source in a spectrum of plane waves

$$\frac{\exp\{ik|\mathbf{r}|\}}{|\mathbf{r}|} = \int_{\mathbf{R}^2} \frac{1}{k_\parallel} \exp\left\{i\mathbf{k}_\perp \cdot \mathbf{r}_\perp + \sqrt{k^2 - |\mathbf{k}_\perp|^2} r_\parallel\right\} d^2\mathbf{k}_\perp. \tag{88}$$

By using the notation

$$\boldsymbol{\kappa}(\mathbf{k}_\perp) = \mathbf{k}_\perp + \hat{\mathbf{n}}\sqrt{k^2 - |\mathbf{k}_\perp|^2}$$

for the "valid" wave vectors (i.e., the wave vectors that belong to the sphere $|\mathbf{k}| = k$, also known as the k-sphere), Weyl's identity is written in simpler form

$$\frac{\exp\{ik|\mathbf{r}|\}}{|\mathbf{r}|} = \int_{\mathbf{R}^2} \frac{1}{k_\parallel} \exp\{i\boldsymbol{\kappa}(\mathbf{k}_\perp) \cdot \mathbf{r}\} \, d^2\mathbf{k}_\perp. \tag{89}$$

Substituting (89) into (87) we obtain successively

$$E_d(\mathbf{r}') = \int_{V_S} \int_{V_{\mathcal{H}}} \int_{\mathbf{R}^2} \int_{\mathbf{R}^2} E_s(\mathbf{r}'')\Delta\epsilon(\mathbf{r})$$
$$\cdot \frac{\exp\{i\boldsymbol{\kappa}(\mathbf{k}_{1,\perp}) \cdot (\mathbf{r}' - \mathbf{r})\}}{k_{1,\parallel}}$$
$$\cdot \frac{\exp\{i\boldsymbol{\kappa}(\mathbf{k}_{2,\perp}) \cdot (\mathbf{r} - \mathbf{r}'')\}}{k_{2,\parallel}}$$
$$d^2\mathbf{k}_{1,\perp} \, d^2\mathbf{k}_{2,\perp} \, d^3\mathbf{r} \, d^3\mathbf{r}'' \tag{90}$$

$$= \int_{\mathbf{R}^2} \int_{\mathbf{R}^2}$$
$$\cdot \left(\int_{V_S} E_s(\mathbf{r}'') \exp\{i\boldsymbol{\kappa}(\mathbf{k}_{2,\perp}) \cdot \mathbf{r}''\} \, d^3\mathbf{r}'' \right)$$
$$\cdot \left(\int_{V_{\mathcal{H}}} \Delta\epsilon(\mathbf{r}) \right.$$
$$\left. \cdot \exp\{i(\boldsymbol{\kappa}(\mathbf{k}_{2,\perp}) - \boldsymbol{\kappa}(\mathbf{k}_{1,\perp})) \cdot \mathbf{r}\} \, d^3\mathbf{r} \right)$$
$$\cdot \exp\{i\boldsymbol{\kappa}(\mathbf{k}_{1,\perp}) \cdot \mathbf{r}'\} \frac{d^2\mathbf{k}_{1,\perp} \, d^2\mathbf{k}_{2,\perp}}{k_{1,\parallel} k_{2,\parallel}} \tag{91}$$

$$= \int_{\mathbf{R}^2} \int_{\mathbf{R}^2} \tilde{E}_s(\boldsymbol{\kappa}(\mathbf{k}_{2,\perp})) \Delta \tilde{\epsilon}(\boldsymbol{\kappa}(\mathbf{k}_{2,\perp}) - \boldsymbol{\kappa}(\mathbf{k}_{1,\perp}))$$

$$\cdot \exp\{i\boldsymbol{\kappa}(\mathbf{k}_{1,\perp}) \cdot \mathbf{r}'\} \frac{d^2 k_{1,\perp} \, d^2 k_{2,\perp}}{k_{1,\parallel} k_{2,\parallel}} \quad (92)$$

$$= \int_{\mathbf{R}^2} \left[\frac{\tilde{E}_s}{k_\parallel} \star \Delta \tilde{\epsilon}^* \right]_{\boldsymbol{\kappa}(\mathbf{k}_{1,\perp})}$$

$$\cdot \exp\{i\boldsymbol{\kappa}(\mathbf{k}_{1,\perp}) \cdot \mathbf{r}'\} \frac{d^2 k_{1,\perp}}{k_{1,\parallel}}. \quad (93)$$

The last statement is equivalent to (73), which proves the assertion.

APPENDIX II
PROOF OF THE CORRELATION PROPERTY FOR STATISTICAL FIELDS

Assuming E_s is one realization of the random process representing the object, we form the correlations according to

$$J_s(\mathbf{r}_1'', \mathbf{r}_2'') = \text{E.V.}\{E_s(\mathbf{r}_1'') E_s^*(\mathbf{r}_2'')\} \quad (94)$$

$$J_d(\mathbf{r}_1', \mathbf{r}_2') = \text{E.V.}\{E_d(\mathbf{r}_1') E_d^*(\mathbf{r}_2')\}. \quad (95)$$

Substituting Born's diffraction formula (87) and (94) into (95), we obtain

$$J_d(\mathbf{r}_1', \mathbf{r}_2') = \int_{V_{\mathcal{H}}} \int_{V_{\mathcal{H}}} \int_{V_S} \int_{V_S}$$

$$\cdot J_s(\mathbf{r}_1'', \mathbf{r}_2'') \Delta \epsilon(\mathbf{r}_1) \Delta \epsilon^*(\mathbf{r}_2)$$

$$\times \frac{\exp\{ikR_1''\}}{R_1''} \frac{\exp\{ikR_2''\}}{R_2''} \frac{\exp\{ikR_1\}}{R_1}$$

$$\cdot \frac{\exp\{ikR_2\}}{R_2} d^3 r_1'' \, d^3 r_2'' \, d^3 r_1 \, d^3 r_2. \quad (96)$$

By using Weyl's identity and proceeding as in Appendix I, we obtain (75) after a long but straightforward calculation.

APPENDIX III
DERIVATION OF THE COHERENT MODE DECOMPOSITION OF THE FIELD DIFFRACTED BY A VOLUME HOLOGRAM

To prove (82) we substitute (76) and (80) into (75). This leads to the expression

$$\tilde{J}_s(\mathbf{k}_1, \mathbf{k}_2) = \sum_{\mathbf{m}} \sum_{\mathbf{j}_1} \sum_{\mathbf{j}_2} \alpha_{\mathbf{m}} \epsilon_{\mathbf{j}_1}^* \epsilon_{\mathbf{j}_2}$$

$$\cdot \iint_{\mathbf{R}^2} \frac{\tilde{\psi}_{\mathbf{m}}^*(\boldsymbol{\kappa}_1'') \tilde{\psi}_{\mathbf{m}}(\boldsymbol{\kappa}_2'')}{k_{1,\parallel} k_{2,\parallel} k_{1,\parallel}'' k_{2,\parallel}''} \tilde{\phi}_{\mathbf{j}_1}^*$$

$$\cdot (\boldsymbol{\kappa}_1'' - \mathbf{k}_1) \tilde{\phi}_{\mathbf{j}_2}(\boldsymbol{\kappa}_2'' - \mathbf{k}_2) \, d^2 k_{1,\perp}'' \, d^2 k_{2,\perp}''. \quad (97)$$

Rearranging the integrals and summations, we rewrite the above expression as

$$\tilde{J}_s(\mathbf{k}_1, \mathbf{k}_2) = \sum_{\mathbf{m}} \alpha_{\mathbf{m}} \left(\sum_{\mathbf{j}_1} \epsilon_{\mathbf{j}_1} \int_{\mathbf{R}^2} \right.$$

$$\left. \cdot \frac{\tilde{\psi}_{\mathbf{m}}(\boldsymbol{\kappa}_1'') \tilde{\phi}_{\mathbf{j}_1}(\boldsymbol{\kappa}_1'' - \mathbf{k}_1)}{k_{1,\parallel} k_{1,\parallel}''} \, d^2 k_{1,\perp}'' \right)^*$$

$$\times \left(\sum_{\mathbf{j}_2} \epsilon_{\mathbf{j}_2} \int_{\mathbf{R}^2} \right.$$

$$\left. \cdot \frac{\tilde{\psi}_{\mathbf{m}}(\boldsymbol{\kappa}_2'') \tilde{\phi}_{\mathbf{j}_2}(\boldsymbol{\kappa}_2'' - \mathbf{k}_2)}{k_{2,\parallel} k_{2,\parallel}''} \, d^2 k_{2,\perp}'' \right) \quad (98)$$

The last expression is equivalent to (82) by using definition (83) for the $\varphi_{\mathbf{m}}$'s. The alternative expression (84) follows from (83) by using Weyl's identity in reverse, in order to go back to the space domain.

REFERENCES

[1] P. J. van Heerden, "Theory of optical information storage in solids," *Appl. Opt.*, vol. 2, no. 4, pp. 393–400, 1963.
[2] A. Ashkin, G. D. Boyd, J. M. Dziedzic, R. G. Smith, A. A. Ballman, and K. Nassau, "Optically-induced refractive index inhomogeneities in LiNbO$_3$," *Appl. Phys. Lett.*, vol. 9, p. 72, 1966.
[3] F. S. Chen, J. T. LaMacchia, and D. B. Fraser, "Holographic storage in lithium niobate," *Appl. Phys. Lett.*, vol. 15, no. 7, pp. 223–225, 1968.
[4] D. Psaltis, "Parallel optical memories," *Byte*, vol. 17, no. 9, p. 179, 1992.
[5] J. F. Heanue, M. C. Bashaw, and L. Hesselink, "Volume holographic storage and retrieval of digital data," *Science*, vol. 265, no. 5173, pp. 749–752, 1994.
[6] D. Psaltis and F. Mok, "Holographic memories," *Sci. Amer.*, vol. 273, no. 5, pp. 70–76, 1995.
[7] Y. S. Abu-Mostafa and D. Psaltis, "Optical neural computers," *Sci. Amer.*, vol. 256, no. 3, pp. 66–73, 1987.
[8] J. Hong, "Applications of photorefractive crystals for optical neural networks," *Opt. Quant. Electr.*, vol. 25, no. 9, pp. S551–S568, 1993.
[9] D. J. Brady, A. G.-S. Chen, and G. Rodriguez, "Volume holographic pulse shaping," *Opt. Lett.*, vol. 17, no. 8, pp. 610–612, 1992.
[10] P.-C. Sun, Y. Fainman, Y. T. Mazurenko, and D. J. Brady, "Space-time processing with photorefractive volume holography," *Proc. SPIE*, vol. 2529, pp. 157–170, 1995.
[11] P.-C. Sun, Y. T. Mazurenko, W. S. C. Chang, P. K. L. Yu, and Y. Fainman, "All-optical parallel-to-serial conversion by holographic spatial-to-temporal frequency encoding," *Opt. Lett.*, vol. 20, no. 16, pp. 1728–1730, 1995.
[12] K. Purchase, D. Brady, G. Smith, S. Roh, M. Osowski, and J. J. Coleman, "Integrated optical pulse shapers for high-bandwidth packet encoding," *Proc. SPIE*, vol. 2613, pp. 43–51, 1996.
[13] D. M. Marom, P.-C. Sun, and Y. Fainman, "Analysis of spatial-temporal converters for all-optical communication links," *Appl. Opt.*, vol. 37, no. 14, pp. 2858–2868, 1998.
[14] G. A. Rakuljic and V. Levya, "Volume holographic narrow-band optical filter," *Opt. Lett.*, vol. 18, no. 6, pp. 459–461, 1993.

[15] C. Mead, "Neuromorphic electronic systems," *Proc. IEEE*, vol. 78, pp. 1629–1636, Sept. 1990.

[16] ——, "Scaling of MOS technology to submicrometer feature sizes," *Analog Int. Circuits Signal Processing*, vol. 6, no. 1, pp. 9–25, 1994.

[17] G. Barbastathis, M. Levene, and D. Psaltis, "Shift multiplexing with spherical reference waves," *Appl. Opt.*, vol. 35, pp. 2403–2417, 1996.

[18] E. N. Leith, A. Kozma, J. Upatnieks, J. Marks, and N. Massey, "Holographic data storage in three-dimensional media," *Appl. Opt.*, vol. 5, no. 8, pp. 1303–1311, 1966.

[19] H. Kogelnik, "Coupled wave theory for thick hologram gratings," *Bell Syst. Tech. J.*, vol. 48, no. 9, pp. 2909–2947, 1969.

[20] D. L. Staebler, J. J. Amodei, and W. Philips, "Multiple storage of thick holograms in LiNbO$_3$," in *Proc. VII Int. Quantum Electronics Conf.*, Montreal, P.Q., Canada, 1972.

[21] D. A. Parthenopoulos and P. M. Rentzepis, "Two-photon volume information storage in doped polymer systems," *J. Appl. Phys.*, vol. 68, no. 11, pp. 5814–5818, 1990.

[22] F. H. Mok, M. C. Tackitt, and H. M. Stoll, "Storage of 500 high-resolution holograms in a LiNbO$_3$ crystal," *Opt. Lett.*, vol. 16, no. 8, pp. 605–607, 1991.

[23] F. H. Mok, "Angle-multiplexed storage of 5000 holograms in lithium niobate," *Opt. Lett.*, vol. 18, no. 11, pp. 915–917, 1991.

[24] F. H. Mok, G. W. Burr, and D. Psaltis, "Angle and space multiplexed random access memory (HRAM)," *Opt. Memory Neural Networks*, vol. 3, no. 2, pp. 119–127, 1994.

[25] R. A. Miller, G. W. Burr, Y.-C. Tai, D. Psaltis, C.-M. Ho, and R. R. Katti, "Electromagnetic MEMS scanning mirrors for holographic data storage," in *Proc. Solid-State Sensor and Actuator Workshop, Tranducer Research Foundation*, Cleveland Heights, OH, 1996, pp. 183–186.

[26] X. An and D. Psaltis, "Experimental characterization of an angle-multiplexed holographic memory," *Opt. Lett.*, vol. 20, no. 18, pp. 1913–1915, 1995.

[27] I. McMichael, W. Christian, D. Pletcher, T. Y. Chang, and J. Hong, "Compact holographic storage demonstrator with rapid access," *Appl. Opt.*, vol. 35, no. 14, pp. 2375–2379, 1996.

[28] D. P. Resler, D. S. Hobbs, R. C. Sharp, L. J. Friedman, and T. A. Dorschner, "High-efficiency liquid-crystal optical phased-array beam steering," *Opt. Lett.*, vol. 21, no. 9, pp. 689–691, 1996.

[29] J.-J. P. Drolet, E. Chuang, G. Barbastathis, and D. Psaltis, "Compact, integrated dynamic holographic memory with refreshed holograms," *Opt. Lett.*, vol. 22, no. 8, pp. 552–554, 1997.

[30] G. A. Rakuljic, V. Levya, and A. Yariv, "Optical data storage by using orthogonal wavelength-multiplexed volume holograms," *Opt. Lett.*, vol. 17, no. 20, pp. 1471–1473, 1992.

[31] S. Yin, H. Zhou, F. Zhao, M. Wen, Y. Zang, J. Zhang, and F. T. S. Yu, "Wavelength-multiplexed holographic storage in a sensitive photorefractive crystal using a visible-light tunable diode-laser," *Opt. Commun.*, vol. 101, nos. 5–6, pp. 317–321, 1993.

[32] C. Denz, G. Pauliat, and G. Roosen, "Volume hologram multiplexing using a deterministic phase encoding method," *Opt. Commun.*, vol. 85, pp. 171–176, 1991.

[33] D. Psaltis, M. Levene, A. Pu, G. Barbastathis, and K. Curtis, "Holographic storage using shift multiplexing," *Opt. Lett.*, vol. 20, no. 7, pp. 782–784, 1995.

[34] M. Mansuripur and G. T. Sincerbox, "Principles and techniques of optical data storage," *Proc. IEEE*, vol. 85, pp. 1780–1796, Nov. 1997.

[35] A. Pu and D. Psaltis, "Holographic 3D disks using shift multiplexing," in *Summaries of Papers Presented CLEO'96*, Baltimore, MD, p. 165.

[36] ——, "Holographic data storage with 100 bits/μm^2 density," in *Proc. Optical Data Storage Topical Meeting*, Tuscon, AZ, 1997, pp. 48–49.

[37] H. Lee, X.-G. Gu, and D. Psaltis, "Volume holographic interconnections with maximal capacity and minimal cross talk," *J. Appl. Phys.*, vol. 65, no. 6, pp. 2191–2194, 1989.

[38] K. Curtis, A. Pu, and D. Psaltis, "Method for holographic storage using peristrophic multiplexing," *Opt. Lett.*, vol. 19, no. 13, pp. 993–994, 1994.

[39] S. Campbell, X. M. Yi, and P. Yeh, "Hybrid sparse-wavelength angle multiplexed optical data storage system," *Opt. Lett.*, vol. 19, no. 24, pp. 2161–2163, 1994.

[40] G. T. Sincerbox, "Holographic storage—The quest for the ideal

[41] M.-P. Bernal, G. W. Burr, H. Coufal, R. K. Grygier, J. A. Hoffnagle, C. M. Jefferson, R. M. McFarlane, R. M. Shelby, G. T. Sincerbox, and G. Wittmann, "Holographic-data-storage materials," *MRS Bull.*, vol. 21, no. 9, pp. 51–60, 1996.

[42] N. V. Kukhtarev, V. B. Markov, S. G. Odulov, M. S. Soskin, and V. L. Vinetskii, "Holographic storage in electrooptic crystals, I. Steady state," *Ferroelect.*, vol. 22, pp. 949–960, 1979.

[43] T. J. Hall, R. Jaura, L. M. Connors, and P. D. Foote, "The photorefractive effect—A review," *Progress Quantum Electron.*, vol. 10, no. 2, pp. 77–145, 1985.

[44] P. Yeh, *Introduction to Photorefractive Nonlinear Optics*. New York: Wiley, 1993.

[45] D. Psaltis, D. Brady, and K. Wagner, "Adaptive optical networks using photorefractive crystals," *Appl. Opt.*, vol. 27, no. 9, pp. 1752–1759, 1988.

[46] F. Mok, G. W. Burr, and D. Psaltis, "A system metric for holographic memory systems," *Opt. Lett.*, vol. 21, no. 12, pp. 896–898, 1996.

[47] G. Barbastathis, J.-J. P. Drolet, E. Chuang, and D. Psaltis, "Compact terabit random-access memory implemented with photorefractive crystals," in *Proc. SPIE Photorefractive Fiber and Crystal Devices: Materials, Optical Properties and Applications III*, San Diego, CA, 1997, pp. 107–122.

[48] N. V. Kukhtarev, V. B. Markov, S. G. Odulov, M. S. Soskin, and V. L. Vinetskii, "Holographic storage in electrooptic crystals, II. Beam coupling—Light amplification," *Ferroelect.*, vol. 22, pp. 961–964, 1979.

[49] P. Yeh, "Two-wave mixing in nonlinear media," *IEEE J. Quantum Electron.*, vol. 25, pp. 484–519, 1989.

[50] A. Yariv, "Phase-conjugate optics and real-time holography," *IEEE J. Quantum Electron.* vol. 14, pp. 650–660, 1978.

[51] J. Feinberg, "Self-pumped continuous-wave phase conjugator using internal reflection," *Opt. Lett.*, vol. 7, no. 10, pp. 486–488, 1982.

[52] D. Z. Anderson and J. Feinberg, "Optical novelty filters," *IEEE J. Quantum Electron.* vol. 25, pp. 635–647, Mar. 1989.

[53] M. Segev, B. Crosignani, and A. Yariv, "Spatial solitons in photorefractive media," *Phys. Rev. Let.*, vol. 68, no. 7, pp. 923–926, 1992.

[54] M.-F. Shih, Z. Chen, M. Mitchell, M. Segev, H. Lee, R. S. Feigelson, and J. P. Wilde, "Waveguides induced by photorefractive screening solitons," *J. Opt. Soc. Amer. B*, vol. 14, no. 11, pp. 3091–3101, 1997.

[55] J. J. Amodei and D. L. Staebler, "Holographic pattern fixing in electro-optic crystals," *Appl. Phys. Lett.*, vol. 18, no. 12, pp. 540–542, 1971.

[56] D. L. Staebler, W. J. Burke, W. Phillips, and J. J. Amodei, "Multiple storage and erasure of fixed holograms in Fe-doped LiNbO$_3$," *Appl. Phys. Lett.*, vol. 26, no. 4, pp. 182–184, 1975.

[57] G. Montemezzani and P. Gunter, "Thermal hologram fixing in pure and doped KNbO$_3$ crystals," *J. Opt. Soc. Amer. B*, vol. 7, no. 12, pp. 2323–2328, 1990.

[58] D. Zhang, Y. Zhang, C. Li, Y. Chen, and Y. Zhu, "Thermal fixing of holographic gratings in BaTiO$_3$," *Appl. Opt.*, vol. 34, no. 23, pp. 5241–5246, 1995.

[59] J. F. Heanue, M. C. Bashaw, A. J. Daiber, R. Snyder, and L. Hesselink, "Digital holographic storage system incorporating thermal fixing in lithium niobate," *Appl. Opt.*, vol. 21, no. 19, pp. 1615–1617, 1996.

[60] A. Y. Liu, M. C. Bashaw, L. Hesselink, M. Lee, and R. S. Feigelson, "Observation and thermal fixing of holographic gratings in lead barium niobate crystal," *Opt. Lett.*, vol. 22, no. 3, pp. 187–189, 1997.

[61] F. Micheron and G. Bismuth, "Electrical control of fixation and erasure of holographic patterns in ferroelectric materials," *Appl. Phys. Lett.*, vol. 20, no. 2, pp. 79–81, 1972.

[62] Y. Qiao, S. Orlov, D. Psaltis, and R. R. Neurgaonkar, "Electrical fixing of photorefractive holograms in (Sr$_{0.75}$Ba$_{0.25}$)Nb$_2$O$_6$," *Opt. Lett.*, vol. 18, no. 12, pp. 1004–1006, 1993.

[63] M. Horowitz, A. Bekker, and B. Fischer, "Image and hologram fixing method with (Sr$_x$Ba$_{1-x}$)Nb$_2$O$_6$ crystals," *Opt. Lett.*, vol. 18, no. 22, pp. 1964–1966, 1993.

[64] R. S. Cudney, J. Fousek, M. Zgonik, P. Günter, M. H. Garrett, and D. Rytz, "Photorefractive and domain gratings in barium titanate," *Appl. Phys. Lett.*, vol. 63, no. 25, pp. 3399–3401, 1993.

material continues," *Opt. Mat.*, vol. 4, nos. 2–3, pp. 370–375, 1995.

[65] J. Ma, T. Chang, J. Hong, R. R. Neurgaonkar, G. Barbastathis, and D. Psaltis, "Electrical fixing of 1,000 angle-multiplexed holograms in SBN:75," *Opt. Lett.*, vol. 22, no. 14, pp. 1116–1118, 1997.

[66] H. C. Külich, "A new approach to read volume holograms at different wavelengths," *Opt. Commun.*, vol. 64, no. 5, pp. 407–411, 1987.

[67] ———, "Reconstructing volume holograms without image field losses," *Appl. Opt.*, vol. 30, no. 20, pp. 2850–2857, 1991.

[68] D. Psaltis, F. Mok, and H.Y.-S. Li, "Nonvolatile storage in photorefractive crystals," *Opt. Lett.*, vol. 19, no. 3, pp. 210–212, 1994.

[69] G. Barbastathis and D. Psaltis, "Shift-multiplexed holographic memory using the two-lambda method," *Opt. Lett.*, vol. 21, no. 6, pp. 429–431, 1996.

[70] E. Chuang and D. Psaltis, "Storage of 1,000 holograms with use of a dual-wavelength method," *Appl. Opt.*, vol. 36, no. 32, pp. 8445–8454, 1997.

[71] D. von der Linde, A. M. Glass, and K. F. Rodgers, "Multiphoton photorefractive processes for optical storage in $LiNbO_3$," *Appl. Phys. Lett.*, vol. 25, no. 3, pp. 155–157, 1974.

[72] ———, "Optical storage using refractive index change induced by two-step excitation," *J. Appl. Phys.*, vol. 47, no. 1, pp. 217–220, 1976.

[73] K. Buse, L. Holtmann, and E. Krätzig, "Activation of $BaTiO_3$ for infrared holographic recording," *Opt. Commun.*, vol. 85, no. 2, pp. 183–186, 1991.

[74] K. Büse, F. Jermann, and E. Krätzig, "Infrared holographic recording in $LiNbO_3$:Cu," *Appl. Phys. A*, vol. 58, no. 3, pp. 191–195, 1994.

[75] K. Büse, A. Adibi, and D. Psaltis, "Non-volatile holographic storage in doubly doped lithium niobate crystals," *Nature*, vol. 393, no. 6686, pp. 665–668, 1998.

[76] D. Psaltis, X. An, G. Barbastathis, A. Adibi, and E. Chuang, "Nonvolatile holographic storage in photorefractive materials," *SPIE Critical Rev.*, vol. CR65, pp. 181–213, 1997.

[77] K. Curtis and D. Psaltis, "Recording of multiple holograms in photopolymer films," *Appl. Opt.*, vol. 31, no. 35, pp. 7425–7428, 1992.

[78] U.-S. Rhee, H. J. Caulfield, and C. S. Vikram, and M. M. Mirsalehi, "Characteristics of the Du Pont photopolymer for angularly multiplexed page-oriented holographic memories," *Opt. Eng.*, vol. 32, no. 8, pp. 1839–1847, 1993.

[79] K. Curtis and D. Psaltis, "Characterization of the Du-Pont photopolymer for 3-dimensional holographic storage," *Appl. Opt.*, vol. 33, no. 23, pp. 5396–5399, 1994.

[80] A. Pu, K. Curtis, and D. Psaltis, "A new method for holographic data storage in photopolymer films," in *Proc. IEEE Nonlinear Optics: Materials, Fundamentals and Applications*, Waikoloa, HI, 1994.

[81] A. Pu and D. Psaltis, "High density recording in photopolymer-based holographic 3D disks," *Appl. Opt.*, vol. 35, no. 14, pp. 2389–2398, 1996.

[82] J. E. Ludman, J. Riccobono, J. Caulfield, J.-M. Fournier, I. Semenova, N. Rienhand, P. R. Hemmer, and S. M. Shahriar, "Porous-matrix holography for nonspatial filtering of lasers," *Proc. SPIE*, vol. 2406, pp. 76–85, 1995.

[83] J. E. Ludman, J. R. Riccobono, N. O. Rienhand, I. V. Semenova, Y. L. Korzinin, S. M. Shahriar, H. J. Caulfield, J.-M. Fournier, and P. Hemmer, "Very thick holographic nonspatial filtering of laser beams," *Opt. Eng.*, vol. 36, no. 6, pp. 76–85, 1997.

[84] G. J. Steckman, I. Solomatine, G. Zhou, and D. Psaltis, "Characterization of phenanthrenequinone-doped poly(methyl methacrylate) for holographic memory," *Opt. Lett.*, vol. 23, no. 16, pp. 1310–1312, 1998.

[85] H.-Y. S. Li, and D. Psaltis, "Three dimensional holographic disks," *Appl. Opt.*, vol. 33, no. 17, pp. 3764–3774, 1994.

[86] B. J. Goertzen and P. A. Mitkas, "Error-correcting code for volume holographic storage of a relational database," *Opt. Lett.*, vol. 20, no. 15, pp. 1655–1657, 1995.

[87] M. A. Neifeld and J. D. Hayes, "Error-correction schemes for volume optical memories," *Appl. Opt.*, vol. 34, no. 35, pp. 8183–8191, 1995.

[88] J. F. Heanue, k. Gurkan, and L. Hesselink, "Signal detection for page-access optical memories with intersymbol interference," *Appl. Opt.*, vol. 35, no. 14, pp. 2431–2438, 1996.

[89] G. W. Burr, J. Ashley, H. Coufal, R. K. Grygier, J. A. Hoffnagle, C. M. Jefferson, and B. Marcus, "Modulation coding for pixel-matched holographic data-storage," *Opt. Lett.*, vol. 20, no. 9, pp. 639–641, 1997.

[90] G. W. Burr, H. Coufal, R. K. Grygier, J. A. Hofnagle, and C. M. Jefferson, "Noise reduction of page-oriented data storage by inverse filtering during recording." *Opt. Lett.*, vol. 5, no. 15, pp. 289–291, 1998.

[91] R. M. Shelby, J. A. Hoffnagle, G. W. Burr, C. M. Jefferson, M.-P. Bernal, H. Coufal, R. K. Grygier, H. Gunther, R. M. McFarlane, and G. T. Sincerbox, "Pixel-matched holographic data storage with megabit pages," *Opt. Lett.*, vol. 22, no. 19, pp. 1509–1511, 1997.

[92] G. Barbastathis, "Intelligent holographic databases," Ph.D. dissertation, California Inst. Technol., Pasadena, 1998.

[93] J. W. Goodman, *Introduction to Fourier Optics*. New York: McGraw–Hill, 1968.

[94] C. Gu, J. Hong, and S. Campbell, "2-d shift-invariant volume holographic correlator," *Opt. Commun.*, vol. 88, nos. 4–6, pp. 309–314, 1992.

[95] F. T. S. Yu and S. Yin, "Bragg diffraction-limited photorefractive crystal-based correlators," *Opt. Eng.*, vol. 34, no. 8, pp. 2224–2231, 1995.

[96] J. R. Goff, "Experimental realization of a multiproduct photorefractive correlation system for temporal signals," *Appl. Opt.*, vol. 36, no. 26, pp. 6627–6635, 1997.

[97] F. T. S. Yu, "Optical neural networks: architecture, design and models," *Progr. Opt.*, vol. 32, pp. 61–144, 1993.

[98] H.-Y. S. Li, Y. Qiao, and D. Psaltis, "Optical network for real-time face recognition," *Appl. Opt.*, vol. 32, no. 26, pp. 5026–5035, 1993.

[99] A. Pu, R. Denkewalter, and D. Psaltis, "Real-time vehicle navigation using a holographic memory," *Opt. Eng.*, vol. 36, no. 10, pp. 2737–2746, 1997.

[100] M. Minsky, "Microscopy apparatus." U.S. Patent. 3 013 467, 1961.

[101] C. J. R. Sheppard and A. Choudhury, "Image formation in the scanning microscope." *Opt. Acta*, vol. 24, pp. 1051–1073, 1977.

[102] C. J. R. Sheppard and C. J. Cogswell, "Three-dimensional image formation in confocal microscopy," *J. Microscopy*, vol. 159, no. 2, pp. 179–194, 1990.

[103] C. J. Cogswell and C. J. R. Sheppard, "Confocal differential interference contrast (DIC) microscopy: Including a theoretical analysis of conventional and confocal DIC imaging," *J. Microscopy*, vol. 165, no. 1, pp. 81–101, 1992.

[104] I. J. Cox, C. J. R. Sheppard, and T. Wilson, "Super-resolution by confocal fluorescence microscopy," *Optik*, vol. 60, pp. 391–396, 1982.

[105] P. T. C. So, T. French, W. M. Yu, K. M. Berland, C. Y. Dong, and E. Gratton, "Time-resolved fluorescence microscopy using two-photon excitation," *Bioimaging*, vol. 3, no. 2, pp. 49–63, 1995.

[106] P. H. Van Cittert, *Physica*, vol. 1, p. 201, 1934.

[107] F. Zernike, *Proc. Phys. Soc.*, vol. 61, p. 158, 1948.

[108] V. Schooneveld, "Image formation from coherence functions in astronomy," in *Proc. IAU Colloquium*, vol. 49, Groningen, 1978.

[109] A. J. Devaney, "The inversion problem for random sources," *J. Math. Phys.*, vol. 20, pp. 1687–1691, 1979.

[110] W. H. Carter and E. Wolf, "Correlation theory of wavefields generated by fluctuating, three-dimensional, primary, scalar sources I. General theory," *Opt. Acta*, vol. 28, pp. 227–244, 1981.

[111] I. J. LaHaie, "Inverse source problem for three-dimensional partially coherent sources and fields," *J. Opt. Soc. Amer. A*, vol. 2, pp. 35–45, 1985.

[112] A. M. Zarubin, "Three-dimensional generalization of the van cittert-zernike theorem to wave and particle scattering," *Opt. Commun.*, vol. 100, nos. 5–6, pp. 491–507, 1992.

[113] J. Rosen and A. Yariv, "Three-dimensional imaging of random radiation sources," *Opt. Lett.*, vol. 21, no. 14, pp. 1011–1013, 1996.

[114] ———, "General theorem of spatial coherence: Application to three-dimensional imaging," *J. Opt. Soc. Amer. A*, vol. 13, no. 10, pp. 2091–2095, 1996.

[115] ———, "Reconstruction of longitudinal distributed incoherent sources," *Opt. Lett.*, vol. 21, no. 22, pp. 1803–1805, 1996.

[116] D. Marks, R. Stack, and D. J. Brady, "3D coherence imaging in the Fresnel domain," *Appl. Opt.*, vol. 38, no. 10, pp. 1332–1342, 1999.

[117] P. K. Rastogi, Ed., *Holographic Interferometry*. Berlin, Germany: Springer-Verlag, 1994.

[118] J. C. Wyant, "Two-wavelength interferometry," *Appl. Opt.*, vol. 10, p. 2113, 1971.

[119] C. Polhemus, "Two-wavelength interferometry," *Appl. Opt.*, vol. 12, no. 9, pp. 2071–2074, 1973.

[120] M. Bertero, P. Brianzi, and E. R. Pike, "Super-resolution in confocal scanning microscopy," *Inv. Probl.*, vol. 3, no. 2, pp. 195–212, 1987.

[121] M. Bertero, P. Boccacci. M. Defrise, C. De Mol, and E. R. Pike, "Super-resolution in confocal scanning microscopy. II. The incoherent case," *Inv. Probl.*, vol. 5, no. 4, pp. 441–461, 1989.

[122] M. Bertero, P. Boccacci, R. E. Davies, and E. R. Pike, "Super-resolution in confocal scanning microscopy. III. The case of circular pupils," *Inv. Probl.*, vol. 7, no. 5, pp. 655–674, 1991.

[123] M. Bertero, P. Boccacci, R. E. Davies, F. Malfanti, E. R. Pike, and J. G. Walker, "Super-resolution in confocal scanning microscopy. IV. Theory of data inversion by the use of optical masks," *Inv. Probl.*, vol. 8, no. 1, pp. 1–23, 1992.

[124] J. G. Walker, E. R. Pike, R. E. Davies, M. R. Young, G. J. Brakenhoff, and M. Bertero, "Superresolving scanning optical microscopy using holographic optical processing," *J. Opt. Soc. Amer. A*, vol. 10, no. 1, pp. 59–64, 1993.

[125] J. Grochmalicki, E. R. Pike, J. G. Walker, M. Bertero, P. Boccacci, and R. E. Davies, "Superresolving masks for incoherent scanning microscopy," *J. Opt. Soc. Amer. A*, vol. 10, no. 5, pp. 1074–1077, 1993.

[126] M. Bertero, P. Boccacci, F. Malfanti, and E. R. Pike, "Super-resolution in confocal scanning microscopy. V. Axial super-resolution in the incoherent case," *Inv. Probl.*, vol. 10, no. 5, pp. 1059–1077, 1994.

[127] D. Psaltis, D. Brady, X. G. Gu, and S. Lin, "Holography in artificial neural networks," *Nature*, vol. 343, no. 6256, pp. 325–330, 1990.

[128] D. Brady and D. Psaltis, "Control of volume holograms," *J. Opt. Soc. Amer. A*, vol. 9, no. 7, pp. 1167–1182, 1992.

[129] G. Barbastathis and D. Psaltis, "Multiplexing methods," in *Holographic Data Storage*, H. Coufal, L. Hesselink, and D. Psaltis, Eds. Berlin, Germany: Springer-Verlag, to be published.

[130] M. Born and E. Wolf, *Principles of Optics*, 6th ed. New York: Pergamon, 1980.

[131] J. W. Goodman, *Statistical Optics*. New York: Wiley, 1985.

[132] L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics*. Cambridge, U.K.: Cambridge Univ. Press, 1995.

George Barbastathis (Member, IEEE) was born in Athens, Greece, in 1971. He received the Diploma in electrical and computer engineering from the National Technical University of Athens in 1993 and the M.Sc. and Ph.D. degrees in electrical engineering from the California Institute of Technology, Pasadena, in 1994 and 1997, respectively. His doctoral dissertation was entitled "Intelligent Holographic Databases."

After postdoctoral work at the University of Illinois, Urbana-Champaign, he joined the faculty at the Massachusetts Institute of Technology, Cambridge, in 1999 as Assistant Professor of Mechanical Engineering. He has extensive experience in the design of holographic memory architectures, interferometric sensors, and learning algorithms. His current research interests are in the applications of optical engineering to machine vision, visual learning, and human–computer interaction, and in optical product design.

Dr. Barbastathis is member of the Optical Society of America and the American Association for the Advancement of Science.

David J. Brady (Member, IEEE) received the B.A. degree in physics and math from Macalester College, St. Paul, MN, in 1985 and the M.S. and Ph.D. degrees in applied physics from the California Institute of Technology, Pasadena, in 1986 and 1990, respectively.

He is an Associate Professor of Electrical and Computer Engineering and a Research Associate Professor in the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign.

His research focuses on optical systems for sensor and communications applications. His main research accomplishments include studies of the control and information capacity of volume holograms, a demonstration of ultrafast three-dimensional (3-D) space-time pulse shaping, and various demonstrations of interferometric and tomographic 3-D imaging.

Limits of Multiplex Imaging

Multiplex sensors and the constant radiance theorem

David J. Brady

Fitzpatrick Center for Photonics and Communication Systems, Department of Electrical and

Computer Engineering, Duke University, Box 90291, Durham, North Carolina 27708

dbrady@duke.edu

We use the coherent mode representation of the cross-spectral density to derive a modal analog of the constant radiance theorem with general applicability to linear optical systems. We use the theorem to consider the relationship between spatial detector geometry and multiplexing capacity.

Optical sensors, such as cameras and grating spectrometers, are usually designed for isomorphic mappings between physical parameters and measured data. With advances in electronic detector arrays and digital processors, however, sensor systems that are deliberately designed with nonisomorphic mappings are increasingly popular. These are termed "multiplex" systems because they measure linear combinations of target data rather than the data itself. Multiplexing has a long history in spectroscopy, as in Fourier and Hadamard transform systems [1, 2], and in x-ray tomography. Motivations for multiplex spectrometry include the "throughput" and "multiplex" advantages. The throughput advantage is that all of the power in the target beam is detected and used to generate the target spectrum. The multiplex advantage is that linear combinations of the target data can increase the mean power per measurement and increase the reconstructed signal to noise ratio in the presence of additive noise. The multiplex advantage is

substantial in the infrared, where thermal noise dominates, but is less compelling in the visible, where shot noise is dominant [3].

There has recently been considerable interest in multiplex techniques for digital imaging. Conventional multiplexing spectroscopy filters the field through a pinhole, slit or fiber to reduce the field to a single spatial mode. Spatial filtering may also be used in scanned imaging systems, as in optical coherence tomography [4]. More commonly, however, imaging systems pass multiple spatial modes. We focus our analysis on systems described by coherent mode decompositions of coherence functions. Multiplex imaging has been demonstrated in a variety of multimode systems [5-10]. Multimode multiplexing is used in imaging systems to capture data for which no isomorphic mapping is possible, as in data radiated by three-dimensional sources, or to improve the efficiency of data capture through target specific mappings of spatial, spectral, polarization and coherence data. Until recently, the nature of measureable sources and the structure of sensor systems was determined by the nature of analog processing in optical systems. Emerging multiplex systems emphasize efficient data transfer over analog data inversion under the assumption that inversion can be digitally implemented after data capture.

Some analyses have directly extended the multiplex advantages of spectroscopy to imaging systems. The conventional analysis assumes, however, that detector noise is independent of the multiplexing scheme. A primary goal of this letter is to show that this assumption cannot hold for multiple spatial mode systems. Multiplexing of multiple modes is constrained by the Second Law of Thermodynamics. The Second Law restricts fan-in and fan-out in optical beams and has broad application to imaging, solar power collection [11-13], and optical interconnections [14].

Second Law restrictions on radiance and fan-in transformations of optical beams have been expressed in many forms and with many names. The most most common form is the constant radiance theorem [15]. These theorems are most commonly derived using ray optics [12], but has also been derived using wave theory [16] and thermodynamic arguments [11]. As suggested by the name, the constant radiance theorem shows that no linear optical system can increase the radiance in transformations between incoherent planes.

In isomorphic imaging systems there is direct analogy between the physical state of the field at significant points and source parameters and one can easily prove the constant radiance theorem by ray arguments. The situation is more complex in multiplex imaging because there need not be any particular physical significance to the field at any point. This letter develops an alternative version of Second Law constraints that allows analysis of systems based on any linear optical transformations and in arbitrary coherence states. We achieve this goal using the modal theory of partial coherence developed by Wolf [17].

We limit our discussion to fields propagating in source-free unbounded homogeneous media, in which case second order coherence functions of the radiant field are completely determined by measures between points on a bounding surface. We can describe the state of these fields and transformations on them using the cross-spectral density $W(\mathbf{r_1},\mathbf{r_2}, \nu)$ and appropriate derivatives for points $\mathbf{r_1}$ and $\mathbf{r_2}$ across a surface bounding the source.

The cross-spectral density is defined as the Fourier transform of the mutual coherence function $\Gamma(\mathbf{r_1},\mathbf{r_2}, \tau)$ [18]. $W(\mathbf{r_1},\mathbf{r_2}, \nu)$ is Hermitian and positive definite in transformations on functions of

$\mathbf{r}_1$ and $\mathbf{r}_2$, by which properties one can show that it can be represented by a coherent mode expansion of the form

$$W(\mathbf{r}_1,\mathbf{r}_2, \nu) = \sum_n \lambda_n(\nu)\phi_n^*(\mathbf{r}_1,\nu)\phi_n(\mathbf{r}_2,\nu) \qquad (1)$$

where $\lambda_n(\nu)$ is real and positive and where the family of functions $\phi_n(\mathbf{r},\nu)$ are orthonormal such that $\int \phi_m^*(\mathbf{r},\nu)\phi_n(\mathbf{r},\nu)\mathrm{d}^2 r = \delta_{mn}$.

In analogy with radiance transformations, we are interested in transformations of the coherent mode distribution for fields propagating between planes. A linear optical system transforms coherent modes defined on an input surface into distributions on an output surface under the impulse response $h(\mathbf{r},\mathbf{r}',\nu)$, where $\mathbf{r}$ and $\mathbf{r}'$ are input and output position vectors, respectively. After propagation through the system the cross-spectral density is

$$W(\mathbf{r}_1',\mathbf{r}_2', \nu) = \sum_n \lambda_n(\nu)\psi_n^*(\mathbf{r}_1',\nu)\psi_n(\mathbf{r}_2',\nu) \qquad (2)$$

where $\psi_n(\mathbf{r}',\nu) = \int \phi_n(\mathbf{r},\nu)h(\mathbf{r},\mathbf{r}',\nu)\mathrm{d}^2 r$. $\mathbf{r}_1'$ and $\mathbf{r}_2'$ correspond to points on a the output surface of the system. The functions $\psi_n(\mathbf{r},\nu)$ are not necessarily orthogonal [19]. The cross-spectral density across the output aperture may be described by a new coherent mode decomposition

$$W(\mathbf{r}_1',\mathbf{r}_2', \nu) = \sum_n \Lambda_n(\nu)\Phi_n^*(\mathbf{r}_1',\nu)\Phi_n(\mathbf{r}_2',\nu) \qquad (3)$$

4

where the functions $\Phi_n(\mathbf{r},\nu)$ are a new set of orthonormal coherent modes and the functions $\Lambda_n(\nu)$ are new eigenvalues. The new coherent modes are complete over the possible states of the field in the output plane, meaning that the states $\psi_n(\mathbf{r},\nu)$ can be expanded as $\psi_n(\mathbf{r},\nu)=\sum_m c_{nm}\Phi_m(\mathbf{r},\nu)$. Using the orthonormality of the coherent modes we know from Eq. (3) that

$$\int W(\mathbf{r}_1',\mathbf{r}_2',\nu)\Phi_m(\mathbf{r}_1',\nu)\Phi_m^*(\mathbf{r}_2',\nu)d^2r_1'd^2r_2'=\Lambda_m(\nu) \qquad (4)$$

Substituting for W in Eq. (4) from Eq. (2) we find

$$\Lambda_m(\nu)=\sum_n |c_{nm}|^2 \lambda_n \qquad (5)$$

Power conservation requirements allow us to constrain the transformation coefficients c_{nm}. Conservation of power on propagation through the system leads to the requirement that $\int \psi_n^*(\mathbf{r},\nu)\psi_n(\mathbf{r},\nu)d^2r \leq 1$, which implies that $\sum_m |c_{nm}|^2 \leq 1$.

$|c_{nm}|^2$ has the properties of a probability distribution. From the weighting implicit to this distribution we see immediately that no linear transformation can increase the maximum mode amplitude, which is to say that $[\Lambda_m]_{max} \leq [\lambda_n]_{max}$. This result is analogous to the constant radiance theorem in that the brightest possible focused spot drawn from the field will be proportional to amplitude of the brightest mode.

If we have no prior knowledge of the original eigenvalues, as in cases for which the values λ_n represent encodable communications, image or memory data, the distribution of λ_n can be taken as a measure of the entropy of the system. Defining the normalized eigenvalue $\tilde{\lambda}_m = \lambda_m / \sum_n \lambda_n$ one constructs the entropy of the system as [20]

$$H = -\sum_n \tilde{\lambda}_n \log \tilde{\lambda}_n \tag{6}$$

One can show by induction that any change in $\tilde{\lambda}_n$ that redistributes the largest $\tilde{\lambda}_n$ into smaller ranges will increase H. Thus, our proof that one cannot increase the largest eigenvalue is simply an expression of the second law of thermodynamics. Of course, linear optical transformations are generally reversible and thus ought not transform the entropy at all. Any transformation that changes the eigenvalue spectrum must be irreversible to satisfy the Second Law. In practical systems irreversibility arises from many factors, including phase loss and nonlinearity on absorption from the field, segmentation of field regions by hard obscurations and slight mechanical instabilities.

In applying our results to multiplex imaging we consider a spatially parallel array of square law spatially and temporally integrating detectors. The state of the ith such detector can be modeled as

$$
\begin{aligned}
m_i &= \int dv \int_{A_i} S(\mathbf{r}_s, v) \kappa(v) d^2 r_s \\
&= \sum_j \int \beta_{ij}(v) \lambda_j(v) dv
\end{aligned}
\tag{7}
$$

where the spatial integral is over the detector area, A_i, and the spectral integral is over all the entire spectrum. $\kappa(v)$ is the spectral efficiency of the detector. $\mathbf{r}_s$ is a position vector on the

detector surface. $S(\mathbf{r}_s, \nu) = W(\mathbf{r}_s, \mathbf{r}_s, \nu)$ is the power spectral density evaluated at $\mathbf{r}_s$. $W(\mathbf{r}_s, \mathbf{r}_s, \nu)$ has been transformed on propagation as discussed above. The power coupling coefficient from the j$^{\text{th}}$ coherent mode in the detector plane to the i$^{\text{th}}$ detector is $\beta_{ij}(\nu) = \kappa(\nu) \int_{A_i} |\phi_j(\mathbf{r}_s, \nu)|^2 d^2 r_s$.

Using the orthonormality of the coherent modes we know that $\sum_{ij} \beta_{ij}(\nu) \leq \kappa(\nu) N$, where N is the number of modes and equality applies if and only if the aggregate detector integration area covers the entire sensor plane. If we assume that the modal distribution is uniform over the sensor plane then we can segment the coupling coefficients to obtain

$$\sum_j \beta_{ij}(\nu) \leq \frac{\kappa(\nu) N A_i}{A_s} \qquad (8)$$

where A_s is the total area of the sensor plane.

In conventional multiplex spectroscopy the input field is single mode and measurements are of the form $m_i = \int \beta_i(\nu) \lambda(\nu) d\nu$. Only power conservation constraints apply to coupling coefficient in these systems, i.e. $\beta_i(\nu) \leq \kappa_{\text{max}}$. Multiplex imaging is complicated by the following factors:

- Source data is encoded both in the mode coefficients $\lambda_i(\nu)$ and in the coupling coefficients $\beta_{ij}(\nu)$. In conventional systems $\beta(\nu)$ is independent of the source state. The relationship between the modes and the coupling coefficients in multiplex systems has the effect of making sensing more difficult and, through the combination the restriction that $\sum_m |c_{nm}|^2 \leq 1$ and Eq. (8), limiting the power on individual sensor elements.

- The range of $\beta_{ij}(v)$ is constrained by Eq. (8). In conventional multiplexing $\beta(v)$ is not strongly coupled to detector size or geometry. As expressed in Eq. (8), multiplexing data from more than one mode is power efficient only if the detector size grows with the number of modes multiplexed. Since detector noise and bandwidth are not independent of detector size, conventional analyses of the multiplex advantage do not necessarily apply to multiplex imaging systems.

Multiplex systems may be subdivided into single mode spectrometers, systems in which the coherent modes are known, such as planar hyperspectral imagers, and systems in which the coherent modes are unknown, such as multidimensional spatio-spectral sensors. The primary advantage of multiplexing in single mode systems is that the net detected power is greater under multiplexing, thereby improving photon efficiency and SNR. The constant radiance theorem suggests that conventional advantages of multiplex sensing are less persuasive in the plane-to-plane imaging case because multiplexing one cannot necessarily increase the power on individual detector elements without increasing the detector area. Depending on detector size and source statistics, however, one may still achieve a multiplexing advantage in planar imaging systems. Comparison of potential advantages with alternative schemes, such as adaptive sensors, is a challenge for future research. Adaptive filtering to discover and match the source modes is the basis of adaptive optical telescopy. Adaptive optical systems correct for relatively weak global distortions of the coherent mode structure. Adaptive filtering for strong distortions and general partially coherent fields has not yet been demonstrated. One might also consider strategy of combining mode powers on absorption through fluorescent mode reduction strategies to

8

address the constant radiance theorem. Entropy constraints in this case will be similar to those in solar power collection [13].

In cases where the coherent modes are either unknown or in which one chooses not to filter on them $\lambda_i(v)$ and $\beta_{ij}(v)$ are unknown. The number of measurements required to resolve these variables depends on the extent of prior knowledge. In most cases one chooses to vary the optical system to generate a full rank linear relationship between the measurements and the source field. Examples of systems that can measure a full rank transformation include direct [9] and indirect [10] coherence measurement systems. How well-conditioned this relationship is depends both on the nature of the source and of the set of optical transformations implemented. The selection of optical transformations or implementation of adaptive systems to achieve well-conditioned transformations is the key challenge in the future design of multiplex imagers.

1. Fateley, W.G., R.M. Hammaker, and R.A. DeVerse, *Modulations used to transmit information in spectrometry and imaging.* Journal of Molecular Structure, 2000. **550-551**: p. 117-22.

2. Harwit, M. and N.J.A. Sloan, *Hadamard transform optics.* 1979, New York: Academic.

3. James, J.F. and R.S. Sternberg, *The design of optical spectrometers.* 1969, London: Chapman and Hall.

4. Huang, D., et al., *Optical coherence tomography.* Science, 1991. **254**: p. 1178-1181.

5. Roddier, F., *Inteferometric Imaging in optical astronomy.* Physics Reports, 1988. **170**(2): p. 97-166.

6. Itoh, K., *Interferometric multispectral imaging*, in *Progress in Optics*, E. Wolf, Editor. 1996, North-Holland: Amsterdam. p. 145-196.

7. E. R. Dowski, J. and W.T. Cathey, *Extended depth of field through wave-front coding.* Applied Optics, 1995. **34**: p. 1859-1866.

8. Rosen, J. and A. Yariv, *Reconstruction of longitudinal distributed incoherent sources.* Optics Letters, 1996. **21**: p. 1803-1806.

9. Marks, D.L., et al., *Visible cone-beam tomography with a lensless interferometric camera.* Science, 1999. **284**(5423): p. 2164-2166.

10. Marks, D.M., R.A. Stack, and D.J. Brady, *Astigmatic coherence sensor for digital imaging.* Optics Letters, 2000. **25**(23): p. 1726-1728.

11. Welford, W.T. and R. Winston, *Optics of nonimaging concentrators.* 1978, San Diego: Academic Press.

12. Welford, W.T. and R. Winston, *High collection nonimaging optics.* 1989, San Diego: Academic Press.

13. Yablonovitch, E., *Thermodynamics of the fluorescent planar concentrator.* Journal of the Optical Society of America, 1980. **70**: p. 1362-1363.

14. Goodman, J., *Fan-in an fan-out with optical interconnections.* Optica Acta, 1985. **32**: p. 1489-1496.

15. Wolfe, W.L., *Introduction to Radiometry.* Tutorial Texts in Optical Engineering, ed. D.C. O'Shea. Vol. TT29. 1998, Bellingham: SPIE Press.

16.   Carter, W.H., *A wave theory for non-imaging concentrators.* Journal of Modern Optics, 1993. **40**: p. 1801-1805.

17.   Wolf, E., *New theory of partial coherence in the space-frequency domain.*

Part I: spectral and cross spectra of steady-state sources. Journal of the Optical Society of America, 1982. **72**: p. 343-351.

18.   Mandel, L. and E. Wolf, *Optical coherence and quantum optics.* 1995, Cambridge: Cambridge University Press.

19.   Wolf, E., *Coherent-mode propagation in spatially band-limited wave fields.* Journal of the Optical Society of America A, 1986. **3**: p. 1920-1924.

20.   O'Neill, E., *Introduction to Statistical Optics.* 1963, Reading: Addison-Wesley.